Herd behavior in consumer inflation expectations - Evidence from the French household survey

Andreas Karpf

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Herd behavior in consumer inflation expectations – Evidence from the French household survey

Andreas KARPF

2013.54

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/bandeau-haut/documents-de-travail/
ISSN : 1955-611X

Herd behavior in consumer inflation expectations - Evidence from the French household survey ☆

Andreas Karpf

Centre d'Économie de la Sorbonne - Université Paris 1 Panthéon-Sorbonne/Paris School of Economics
Maison des Sciences Économiques, 106 - 112 boulevard de L'Hôpital, 75647 Paris cedex 13, France
phone: +33-6-99-31-79-12

Abstract

This article investigates whether the formation of individual inflation expectations is biased towards a consensus and is thus subject to some kind of herding behavior. Basing on the traditional Carlson-Parkin approach to quantify qualitative survey expectations and its extension by Kaiser and Spitz (2002) in an ordered probit context, a method to gain individual level inflation expectations is proposed using a Markov chain Monte Carlo Hierarchical Bayesian estimation method. This method is applied to micro survey data about inflation expectations of households from the monthly French household survey "Enquête mensuelle de conjoncture auprès des ménages - ECAMME" (January 2004 to December 2012). Finally a non-parametric test for herding behavior (Bernardt et al., 2006) is conducted on the cohort-level expectation estimates, showing that the expectation formation is not subject to a bias towards the expectation consensus. In contrast, it exhibits a strong anti-herding tendency which is consistent with the findings of other studies (Rülke and Tillmann, 2011).

Keywords: herd behavior, inflation, rational expectations
JEL: E31, E37, D12, D82, D84

☆WORKING PAPER: last revision 2013-07-04
Email address: andreas.karpf@malix.univ-paris1.fr (Andreas Karpf)

Contents

1. Introduction

Assumptions about expectations regarding inflation are exceedingly relevant for economic theory as well as policy making. Consumer surveys measuring households' perceptions and expectations regarding the evolution of prices thus have developed into important supplementary tools for monetary authorities and a vivid field of research. The latter is foremost motivated, besides the fact that inflation is an important economic variable directly impacting the welfare of households, by the discussion if and to what degree inflation is fully anticipated and thus if expectations are rational or unbiased. The falsification or verification of several economic theories as for example the well known Phillips curve (1958), heavily base on this question. The famous critique of Robert Lucas (1976) and Milton Friedman (1968) of the aforementioned theory, which was rather an empirical finding by William Phillips, attach to the question of expectational rationality: The goal of lowering unemployment under its natural rate with the help of for example monetary policy would at least in the medium run be offset by agents rationally anticipating a higher inflation rate in the future and comprising it for example in their wage bargaining.

Given the importance of expectation formation and the question if it is rational or not for economic theory, it is not surprising that a huge body of literature, bot theoretical and empirical, emerged around this topic. Theoretical work on rational expectations within economics started with seminal works by Muth (1961), Sargent et al. (1973) and, cited already above, Lucas (1976). These works were so influential that today most economic models comprise rationality of expectations or unbiasedness of expectations within the core of their model assumptions. This however does not imply that this assumption is not disputed. There are several more recent works which empirically test this hypothesis, some of them confirming the hypothesis of rational expectations, as Thomas (1999) or Ang et al. (2007), others, at least partially, rejecting it, like Mehra (2002), Mankiw and Reis (2002), Roberts (1997) or Baghestani (2009).

Using microlevel data from the monthly French household survey (Enquête mensuelle de conjoncture auprès des ménages - ECAMME)[1] this paper addresses the problem field of rationality or unbiasedness of consumer (household) expectations from a different perspective which by now got fairly little attention. It is investigated if some kind or *herd— or flocking— behavior* is identifiable within the expectation formation of consumers/households. Herding behavior in this context is defined as a bias towards the consensus of expectations which is assumed to be the mean of all prior expectations within a period. This issue you will be discussed in detail later on. The structure of the study is as follows: The traditional probabilistic method, see Theil (1952) and Carlson and Parkin (1975), to quantify survey expectations is extended in a hierarchical Bayesian ordered probit framework to gain individual/cohort-level inflation expectations. Applying a non-parametric test by Bernardt et al. (2006) of herding behavior to the quantified cohort-level inflation expectation estimates finally allows to investigate if consumer expectations of inflation are solely based on the individual assessment (anti-herding) or if they are biased in the direction of a general sentiment or consensus (herding).

The only work, the author is aware of that investigates herding behavior in the context of surveys is the work by Franke (2007). This paper develops a microfounded model of herding in which agents can switch between two states, optimistic and pessimistic. By means of business survey data from the German ifo and ZEW survey Franke shows that there is an empirically significant co-movement of agents in terms of transition probabilities between the two states. The paper at hand is different in two ways: First, it addresses herding behavior of consumers with regard to inflation expectations instead of business sentiment, second, the research question of herding is addressed in a quantitative instead of a qualitative manner as in the paper cited above, thus seeks to answer the question if respondents are biased in the direction of a quantitative consensus.

2. The Data - Enquête mensuelle de conjoncture auprès des ménages

The Enquête mensuelle de conjoncture auprès des ménages, in the remainder abbreviated by the official acronym ECAMME is a monthly survey conducted by the French statistical office the *Institut National de la Statistique et des Études Économiques* (short INSEE) since 1987, replacing the ancient household survey which took place three times

[1]This survey has already been used by other authors to investigate the issue of rational expectations as in example Gardes and Madre (1991); Gardes et al. (2000) or Gourieroux and Pradel (1986).

a year. The ECAMME is part of the *Harmonised EU Programme of Business and Consumer Surveys of the European Commission* which has the goal to standardize survey based economic research within the European Union. The ECAMME is conducted via telephone interviews with approximately 3300 households per month (until 2006 with the exception of August), which are randomly selected from the official French telephone register. The survey collects information about the financial situation, employment and the standard of living of the interviewed households as well as their perceptions and expectations regarding various economic variables. In the context of this paper question 5 and 6 within ECAMME are of importance, which ask for the households perceptions and expectations with regard to past and future consumer price developments:

(Q5) Do you think that prices in the last twelve months have ... (*Trouvez-vous que, au cours des douze derniers mois, les prix ont...*)

- increased strongly (*fortement augmenté*)
- increased moderately (*modérément augmenté*)
- stagnated (*stagné*)
- decreased (*diminué*)

(Q6) In comparison with the last twelve months how do you think the evolution price will be in the next twelve months ... (Par rapport aux douze derniers mois, quelle sera à votre avis l'évolution des prix au cours des douze prochains mois ...)

- prices will increase with a higher rate (*elle va être plus rapide*)
- prices will increase with the same rate (*elle va se poursuivre au même rythme*)
- prices will increase with a smaller rate (*elle va être moins rapide*)
- prices will stay the same (*les prix vont rester stationnaires*)
- prices will go down (*les prix vont diminuer*)

For the here conducted research micro data from January 2004 until December 2012 was available, supplied by Réseau Quetelet as a distributor for INSEE.[2] After sorting out non responses, especially in Q5 and Q6, and flawed data, this corresponds to all in all 185,945 observations or approximately 1788 usable interviews per month. The data contains a wide variety of socio-economic information, for example household size, level of education of the head of the household as well as his/her companion, employment status of the head of the household as well as his/her companion, income quartile, age, region, the number of children, the number of persons living in the household et cetera.

ECAMME covers a wide range of the French society: The average participant in the available dataset is however 55.4 years old (st. dev. 16.58), has 0.4 (st.dev. 0.81)

[2]The reader is referred to section 10.1 for detailed data references.

children and lives in a household with 2.4 persons (st.dev. 1.3). Of the individuals in the data set 23 % finished primary and 27.5 % finished secondary education. 20.2% had completed a post-secondary school and 29.2% held a university degree. For some descriptive statistics of the available ECAMME dataset the reader is referred to Table 1 and Table 2.

		age		children		hh.size	
sex	education	mean	st. dev.	mean	st. dev.	mean	st. dev.
male	primary	68.14	12.47	0.09	0.44	1.98	1.04
	secondary	55.90	15.19	0.32	0.74	2.42	1.22
	post secondary	52.52	14.38	0.41	0.80	2.51	1.23
	tertiary	49.31	15.90	0.51	0.89	2.59	1.31
female	primary	68.99	12.44	0.08	0.43	1.81	1.04
	secondary	55.42	15.92	0.42	0.82	2.48	1.35
	post secondary	50.74	14.44	0.54	0.90	2.71	1.33
	tertiary	45.99	14.34	0.68	0.97	2.76	1.38

Table 1: Descriptive Statistics - ECAMME data set (Jan 2004 - Dec 2012)

sex	education	income			
		1st quart.	2nd quart.	3rd quart.	4th quart.
male	primary	42.92	33.59	17.25	6.25
	secondary	17.83	26.15	30.97	25.06
	post secondary	17.4	26.06	35.01	21.53
	tertiary	7.48	11.94	24.17	56.41
female	primary	54.22	30.55	11.72	3.51
	secondary	24.08	28.71	28.87	18.35
	post secondary	22.2	27.53	32.32	17.95
	tertiary	9.88	16.27	26.41	47.44
male		19.38	23.06	27.04	30.53
female		27.36	25.52	24.53	22.58

Table 2: Descriptive Statistics - Income - ECAMME data set (Jan 2004 - Dec 2012)

3. A simple non-parametric test for herding

3.1. The idea

In this section a simple non-parametric test for herding is introduced which was originally developed by Bernardt et al. (2006) to test for a potential biasedness of professional forecasters. It is then shown how this test could be applied to consumer survey data.

It is assumed that consumers intrinsically form expectations over future developments in example of prices in a similar way professional forecasters do, by taking into account every disposable information or evidence (this means their own daily consumption experience, communication with other people, the consumption of media et cetera). The difference of course is that consumers, uncomfortable with economic measures, might have difficulties in quantifying inflation within the next months. This problem is addressed in consumer sentiment or household surveys by asking for qualitative tendencies rather than for exact numbers. Evidence shows that the aggregation of such sentiments delivers a pretty precise picture of the future evolution of prices (Ludvigson, 2004; Mourougane and Roma, 2003; Howrey, 2001; Vuchelen, 2004; Vuchelen and Praet, 1984). The problem of quantifying consumer expectations and thus how to gain quantitative forecasts from qualitative consumer expectations collected by surveys (similar to earning forecasts by analysts) on an individual/cohort-level will be addressed in the next section.

For the sake of clarity, the terminology of the literature of finance is adopted: A forecast in this sense is a quantified formulation of expectations over the future development of an economic variable, here inflation $\pi^e_{t,t+1}$. A consensus forecast $\bar{\pi}^e$ is understood as the aggregated and quantified expectation of a reference group for example other individuals which formulated their expectations at an earlier point in time (later on, the mean of all prior forecasts for the same target value is used as the consensus). A forecast $\pi^e_{t,t+1}$ at time t for inflation π_{t+1} at time $t+1$ is regarded as unbiased if, given all available informations, it equals the median of all posteriors $\hat{\pi}^e_{t,t+1}$, this means $\pi^e_{t,t+1} = \hat{\pi}^e_{t,t+1}$. Thus, if forecasts are unbiased, there is no reason to assume that they generally tend to be higher or lower than the realized value of the forecasted quantity. In this sense the probability, given the available information set, that a forecast exceeds or falls short of the realized value π_{t+1} can be assumed to be equally 0.5: $P(\pi^e_{t,t+1} < \pi_{t+1}) = P(\pi^e_{t,t+1} > \pi_{t+1}) = 0.5$. If a forecast is however biased it can be assumed that it deviates from the median of posteriors. Therefore the probabilities that the realized values of the forecasted quantities will be above or below the forecast also change. In terms of herding, a bias will be one towards the extant consensus of a reference group (the mean of prior expectations/forecasts with regard to the same variable of [all] other individuals). If an agent herds and his forecast lies above the consensus then the probability that his forecast will be too low is more than one half. Vice versa the probability that a forecast will exceed the realized value given a bias towards the consensus where the forecast lies below the consensus is equally more than one half. Thus, seen from the opposite perspective and more formal: If the agent herds toward the consensus $\bar{\pi}^e_{t,t+1}$ and his posterior $\hat{\pi}^e_{t,t+1}$ is above the consensus he will choose a forecast $\pi^e_{t,t+1} \in \{\bar{\pi}^e_{t,t+1}, \hat{\pi}^e_{t,t+1}\}$. So if $\pi^e_{t,t+1} > \bar{\pi}^e_{t,t+1}$, it will exceed the realized value with probability less than one half, as $\pi^e_{t,t+1} < \hat{\pi}^e_{t,t+1}$ and $P(\pi_{t+1} < \pi^e_{t,t+1}) < P(\pi_{t+1} < \hat{\pi}_{t,t+1}) = \frac{1}{2}$. Herding can be assumed if the two following conditional probabilities fulfill the following conditions:

$$P(\pi_{t+1} < \pi^e_{t,t+1} | \bar{\pi}^e_{t,t+1} < \pi^e_{t,t+1}, \pi^e_{t,t+1} \neq \pi_{t+1}) < \frac{1}{2} \tag{1}$$

$$P(\pi_{t+1} > \pi^e_{t,t+1} | \bar{\pi}^e_{t,t+1} > \pi^e_{t,t+1}, \pi^e_{t,t+1} \neq \pi_{t+1}) < \frac{1}{2} \tag{2}$$

Anti-herding on the other hand, thus a bias away from the consensus forecast, is fulfilled if:

$$P(\pi_{t+1} < \pi^e_{t,t+1} | \bar{\pi}^e_{t,t+1} < \pi^e_{t,t+1}, \pi^e_{t,t+1} \neq \pi_{t+1}) > \frac{1}{2} \tag{3}$$

$$P(\pi_{t+1} > \pi^e_{t,t+1} | \bar{\pi}^e_{t,t+1} > \pi^e_{t,t+1}, \pi^e_{t,t+1} \neq \pi_{t+1}) > \frac{1}{2} \tag{4}$$

3.2. The Test Statistics

With regard to the idea presented in Section 2.1 Bernardt et al. (2006) construct the following test statistics which is also used here. The conditioning events z_t^+, if $\pi^e_{t,t+1} > \bar{\pi}^e_{t,t+1}$, and z_t^-, if $\pi^e_{t,t+1} < \bar{\pi}^e_{t,t+1}$, are defined. According to this the indicator functions,

$$\gamma_t^+ = 1 \quad if \quad z_t^+ \quad \text{otherwise} \quad \gamma_t^+ = 0 \tag{5}$$

$$\gamma_t^- = 1 \quad if \quad z_t^- \quad \text{otherwise} \quad \gamma_t^- = 0 \tag{6}$$

are constructed. The variables

$$\delta_t^+ = 1 \quad if \quad z_t^+ \quad AND \quad \pi_t^e > \pi_t \quad \text{otherwise} \quad \delta_t^+ = 0 \tag{7}$$

$$\delta_t^- = 1 \quad if \quad z_t^- \quad AND \quad \pi_t^e < \pi_t \quad \text{otherwise} \quad \delta_t^- = 0 \tag{8}$$

indicate overshooting and undershooting with regard to the realized value.

The mean of both conditional probabilities from above measures if the forecasts overshoot/undershoot the realized variable in the same direction in which they overshoot/undershoot the consensus forecast.

$$S(z_t^-, z_t^+) = \frac{1}{2}\left[\frac{\sum_t \delta_t^+}{\sum_t \gamma_t^+} + \frac{\sum_t \delta_t^-}{\sum_t \gamma_t^-}\right] \tag{9}$$

$S(z_t^-, z_t^+) < \frac{1}{2}$ indicates a bias to the consensus (herding) while $S(z_t^-, z_t^+) > \frac{1}{2}$ indicates a bias away from the consensus (anti-herding). A derivation of the second central moment of the test statistics as well as a discussion of possible robustness issues can be found in Appendix A and B respectively (along Bernardt et al. (2006)).

4. Quantifying Inflation Expectations

Unlike to the forecasts of professional analysts, the method described in the previous section cannot directly be applied to consumer expectation data collected by surveys. Quantified consumer expectations with regard to inflation are not, or just rarely, available. Interviewees participating in a consumer survey like ECAMME might be unfamiliar to give a concrete quantitative answer how for example prices will evolve in the upcoming twelve months. Therefore consumer surveys normally ask for tendencies rather than precise numbers, by proposing qualitative response options. Before the above described test can be applied, estimation techniques have to be used to transform the qualitative answers of survey participants into quantitative forecasts.[3]

[3]Since 2004 ECAMME contains questions for the quantitative perceptions and expectations of inflation. Analyzing this data shows that respondents in aggregate have a good intuition of price developments as far as tendencies are concerned, but are rather bad at giving quantitative estimates. The concerned variables are full of outliers and frequently state totally exaggerated values.The monthly averages of these quantitative estimations by respondents lie systematically several percentage points above the actual inflation rate. A potential use of these variables is however briefly discussed in Section 7.2.

Roughly spoken there are two different approaches to quantify qualitative survey data: The regression approach which roots can be tracked back to Anderson (1952), Pesaran (1985, 1989) as well as Pesaran and Weale (2006) on the one hand, and the probability approach which was initially developed by Theil (1952) and Carlson and Parkin (1975) respectively and thus is often denominated by the latter as the *Carlson-Parkin* approach on the other hand. In the paper at hand a modification of the probability or Carlson-Parkin approach will be used: Along the paper by Kaiser and Spitz (2002) the Carlson-Parkin method will be interpreted in the context of an ordered probit/logit estimation. As will be seen in the following discussion, such a modification of the probability approach provides two crucial advantages for the here envisioned task when compared to the regression approach:

1. it bases to a lesser extent on strict assumptions,
2. the various socio-demographic information available in the micro data of the ECAMME survey can be exploited to derive individual level-inflation expectations when the probability approach is interpreted as an ordered probit/logit model and estimated in a hierarchical bayesian framework.

Corresponding to questions 5 and 6 in ECAMME (see Section 2), an expectation horizon of twelve months is used in the notation within this section.[4] This was of course taken into account when the estimations where done and the test statistic was applied respectively.

4.1. Regression approach

The regression approach in its baseline setting, as presented by Pesaran and Weale (2006), assumes that there are only two different answering options in a survey regarding perceived and expected inflation and that each of the N participants j had a specific expected inflation rate $\pi^e_{j,t,t+1}$ in mind when surveyed at time t. If one would then group the participants as U_{t+1} and D_{t+1} depending on whether they expected rising (denoted as $+$) or falling prices (denoted as $-$), one could write:

$$\tilde{\pi}^e_{t,t+1} = \sum_{j \in U_{t+1}} w^+_{j,t+1} \pi^{e+}_{j,t,t+1} + \sum_{j \in D_{t+1}} w^-_{j,t+1} \pi^{e-}_{j,t,t+1} \tag{10}$$

Unfortunately specific values of $\pi^e_{j,t,t+1}$ are not available for household surveys. Pesaran therefore supposes that inflation expectations of households, may they indicate a lower or higher rate of inflation, fluctuate around a fixed moment with independently distributed error terms ϵ_i for each individual with mean 0 and variance σ^2. Under this assumption the expected inflation rates can be expressed as:

$$\pi^{e+}_{j,t,t+1} = \alpha + \epsilon_{j\alpha} \tag{11}$$

$$\pi^{e-}_{j,t,t+1} = -\beta + \epsilon_{j\beta} \tag{12}$$

[4]Before 2004, when the survey was adjusted to the standard of the harmonized European consumer surveys program, participants where asked for a perception/expectation horizon of six months. Therefore not the whole monthly data set of ECAMME could be used.

with $\alpha, \beta > 0$. If the variances σ^2_{alpha} and σ^2_{beta} are sufficiently small and follow appropriate shaped distributions in order that $\pi^{e+}_{j,t+1} > 0$ and $\pi^{e-}_{j,t+1} < 0$ respectively $\forall j, t$ then one can rewrite (10) to:

$$\tilde{\pi}^e_{t,t+1} \approx \alpha \sum_{j \in U_{t+1}} w^+_{j,t} - \beta \sum_{j \in D_{t+1}} w^-_{j,t+1} \tag{13}$$

$$\tilde{\pi}^e_{t,t+1} \approx \alpha U^e_{t+1} - \beta D^e_{t+1} \tag{14}$$

The crucial and at the same time very strong assumption by Pesaran is that inflation perceptions are formed in the same manner as inflation expectations: parameters α and β can be estimated by means of appropriate regression techniques with π_t on the fractions U_t and D_t regarding inflation perceptions and then used to calculate a time series of quantitative values for $\pi^e_{t,t+1}$. This means in example for a linear scenario and the here used twelve month horizon in both directions that $\pi_t = \alpha U_t + \beta D_t + \epsilon_t$ is estimated for the actual inflation rate (which in the standard monthly form gives the evolution of consumer prices between $t - 12$ and t), where the parameters $\hat{\alpha}$ and $\hat{\beta}$ are then used to calculate the expected inflation rate by $\hat{\pi}_{t,t+12} = \hat{\alpha} U_t + \hat{\beta} D_t$ (Pesaran and Weale, 2006).

There are various extensions of the baseline model. A modification proposed by Pesaran and Weale (2006) comprises an adjustment to an assumed asymmetry between the perception of rising and falling prices, where it is supposed that the former outweighs the latter: $\pi_t = (\alpha U_t + \beta D_t)/(1 - \lambda U_t) + \epsilon_t$. Another modification includes AR(2) autoregressive errors from the OLS version of the model in order to correct for autocorrelation of the error terms: $\pi_t = (\alpha U_t + \beta D_t + \phi_1 \hat{\epsilon}_{t-1} + \phi_2 \hat{\epsilon}_{t-2})/(1 - \lambda U_t) + \epsilon_t$. Along the reviewed literature this procedure is regarded as disputable (Curto Millet, 2006; Nardo, 2003), but might serve to gain a good fit between estimated expected inflation and actual inflation by accounting for some kind of eventual error correction algorithm applied by agents when uttering their expectations and perceptions. Accounting for the equational form both models are estimated by means of non-linear least squares estimations.

The regression approach however has two characteristics which renders it problematic for the here envisioned task: First, perceptions of price changes are assumed to be unbiased; second, it is assumed that inflation perceptions are formed along the same mechanisms as inflation expectations; third, and more important for the envisioned task, it does not allow to control for heterogeneity within participants.

4.2. The Carlson-Parkin Approach

The Carlson-Parkin or probability approach chooses a different way to quantify qualitative expectation data from a survey: Namely it assumes that the fraction of each answering option corresponds to a maximum likelihood estimate in the context of the aggregate density function with regard to inflation expectations (Forsells and Kenny, 2002). Simply put, the perceived inflation rate is anchored to an assumed probability distribution function using the qualitative survey data of inflation expectations. To demonstrate this approach it is referred to Berk (1999) who also discusses a survey with five answering categories like it is the case in the forward looking inflation question in the ECAMME (Q6).

The perceived inflation rate is anchored to an assumed probability distribution via thresholds: It is for example assumed that the expected inflation follows a Gaussian distribution.[5] From the data, by the fraction respondents chose each response category, so called threshold values are calculated. At these cut-off points people change one answering option for another. These points are then scaled to the perceived inflation rate. The perceived inflation is the assumed rate of the increase of prices people have in mind when choosing one of the answering categories in the survey with regard to the price developments in the upcoming twelve months. In this respect two intervals are important in the baseline Carlson-Parkin approach:

- δ_t is an interval in which consumers perceive no change in prices. This interval is called the *indifference limen.*
- μ_t signifies an interval around the perceived rate of inflation $\pi^p_{j,t}$ above/below which consumers experience or better expect an increasing/decreasing rate of inflation $\pi^e_{j,t,t+1}$.

It is important to note that both thresholds are symmetric, a shortcoming which will be addressed later on by the ordered probit interpretation of the Carlson-Parkin method. In the baseline model it is further assumed that the perceived rate of inflation $\pi^p_{j,t}$ is equal for all individuals j and can thus be denoted as π^p_t. In the simplest scenario it is assumed that the perceived inflation rate is the latest published inflation rate. This is of course disputable, since individuals may have a different consumption attitude and might thus perceive price changes differently. Since ECAMME, like all standardized European household surveys, also contains a question how respondents perceived inflation in the last twelve months (Q5), it is easy to also estimate quantified values of the perceived inflation rate by the Carlson-Parkin method.

The threshold model with δ_t and μ_t can be written as (Maag, 2009):

$$
\begin{aligned}
\pi^e_{j,t} < -\delta_t &: \text{the prices will decrease } (S_1)\\
-\delta_t \leq \pi^e_{j,t} < \delta_t &: \text{the prices will stay the same } (S_2)\\
\delta_t \leq \pi^e_{j,t} < \pi^p_t - \mu_t &: \text{the prices will rise at a lower rate } (S_3)\\
\pi^p_t - \mu_t \leq \pi^e_{j,t} < \pi^p_t + \mu_t &: \text{prices will rise at the same rate } (S_4)\\
\pi^e_{j,t} < \pi^p_t + \mu_t &: \text{prices will rise at a faster rate } (S_5)
\end{aligned}
$$

The probabilities for these events can be easily estimated by the response shares. If it is assumed that the expected inflation rate follows a normal distribution $\pi^e_{j,t} \sim N(\pi^e_t, (\sigma^e_t)^2)$ and that $\Phi()$ is the c.d.f of the normal distribution, the threshold model from above can

[5]This is the standard assumption but was for example criticized by Maddala (1991). Berk (1999) for example tests the Carlson-Parkin approach under different distributional assumptions.

be rewritten in terms of probabilities:

$$s_t^1 = P(\pi_{j,t}^e < -\delta_t) = \Phi\left(\frac{-\delta_t - \pi_t^e}{\sigma_t}\right) \tag{15}$$

$$s_t^2 = P(-\delta_t \leq \pi_{j,t}^e < \delta_t) = \Phi\left(\frac{\delta_t - \pi_t^e}{\sigma_t}\right) - \Phi\left(\frac{-\delta_t - \pi_t^e}{\sigma_t}\right) \tag{16}$$

$$s_t^3 = P(\delta_t \leq \pi_{j,t}^e < \pi_t^p - \mu_t) = \Phi\left(\frac{\pi_t^p - \mu_t - \pi_t^e}{\sigma_t}\right) - \Phi\left(\frac{\delta_t - \pi_t^e}{\sigma_t}\right) \tag{17}$$

$$s_t^4 = P(\pi_t^p - \mu_t \leq \pi_{j,t}^e < \pi_t^p + \mu_t) = \Phi\left(\frac{\pi_t^p + \mu_t - \pi_t^e}{\sigma_t}\right) - \Phi\left(\frac{\pi_t^p - \mu_t - \pi_t^e}{\sigma_t}\right) \tag{18}$$

$$s_t^5 = P(\pi_{j,t}^e \geq \pi_t^p + \mu_t) = 1 - \Phi\left(\frac{\pi_t^p + \mu_t - \pi_t^e}{\sigma_t}\right) \tag{19}$$

Using the inverse cumulative distribution allows to rewrite this system of equations which makes it solvable for the unknowns π_t^e, σ_t, δ_t and μ_t. π_t^p is assumed to be given by the most recently published inflation rate in the baseline model.

$$\begin{aligned}
G_t^1 &= \Phi^{-1}(s_t^1) = \frac{-\delta_t - \pi_t^e}{\sigma_t} \\
G_t^2 &= \Phi^{-1}(1 - s_t^5 - s_t^4 - s_t^3 - s_t^2) = \frac{-\delta_t - \pi_t^e}{\sigma_t} \\
G_t^3 &= \Phi^{-1}(1 - s_t^5 - s_t^4 - s_t^3) = \frac{\delta_t - \pi_t^e}{\sigma_t} \\
G_t^4 &= \Phi^{-1}(1 - s_t^5 - s_t^4) = \frac{\pi_t^p - \mu_t - \pi_t^e}{\sigma_t} \\
G_t^5 &= \Phi^{-1}(1 - s_t^5) = \frac{\pi_t^p + \mu_t - \pi_t^e}{\sigma_t}
\end{aligned}$$

The unknown variables can easily be found by combining the equations from above. The mean expected inflation rate would in example be:

$$\pi_t^e = \pi_t^p \frac{G_t^2 + G_t^3}{G_t^2 + G_t^3 - G_t^4 - G_t^5}$$

4.3. Quantification with Ordered Probit

Reconsidering the approach by Carlson-Parkin it can be formulated as a threshold model which could equivalently be estimated by the use of an ordered probit regression (Kaiser and Spitz, 2002). Again it is assumed that respondent j within a survey base his decision which answer on the scale to choose on a subliminal threshold ranking. The estimation of the model by the use of ordered probit however allows for asymmetric thresholds (μ_1, μ_2, μ_3, μ_4). This seems more appropriate as the decrease/increase of prices is likely to be perceived with a different sensitivity. The expected change between perceived and future inflation (the forward looking question Q6 asks for the change of inflation, while the backward looking question Q5 asks for the change of prices; this will be discussed in more detail later on) is now denoted as $\Delta\pi_t^e = \pi_t^e - \pi_t^p$.

$$\begin{aligned}
&\pi_t^e - \pi_t^p < \mu_t^1 \\
&\mu_t^1 \leq \pi_t^e - \pi_t^p < \mu_t^2 \\
&\mu_t^2 \leq \pi_t^e - \pi_t^p < \mu_t^3 \\
&\mu_t^3 \leq \pi_t^e - \pi_t^p < \mu_t^4 \\
&\pi_t^e - \pi_t^p \geq \mu_t^4
\end{aligned}$$

This implies along before: [6]

$$\begin{aligned}
&P(S_1) = P(\Delta\pi^e \leq \mu_1) = P(0 \leq -\Delta\pi^e + \mu_1) \\
&P(S_2) = P(\Delta\pi^e \leq \mu_2) - P(\Delta\pi^e \leq \mu_1) = P(0 \leq -\Delta\pi^e + \mu_2) - P(0 \leq -\Delta\pi^e + \mu_1) \\
&P(S_3) = P(\Delta\pi^e \leq \mu_3) - P(\Delta\pi^e \leq \mu_2) = P(0 \leq -\Delta\pi^e + \mu_3) - P(0 \leq -\Delta\pi^e + \mu_2) \\
&P(S_4) = P(\Delta\pi^e \leq \mu_4) - P(\Delta\pi^e \leq \mu_3) = P(0 \leq -\Delta\pi^e + \mu_4) - P(0 \leq -\Delta\pi^e + \mu_3) \\
&P(S_5) = 1 - P(\Delta\pi^e \leq \mu_4) = 1 - P(0 \leq -\Delta\pi^e + \mu_4)
\end{aligned}$$

As before, it is assumed that the expected inflation rate follows a normal distribution. $\Phi()$ is the cumulative distribution function.

$$\begin{aligned}
&S_1 = P(\Delta\pi^e < \mu_1) = \Phi(\frac{-\Delta\pi^e + \mu_1}{\sigma}) \\
&S_2 = P(\mu_1 \leq \Delta\pi^e < \mu_2) = \Phi(\frac{-\Delta\pi^e + \mu_2}{\sigma}) - \Phi(\frac{-\Delta\pi^e + \mu_1}{\sigma}) \\
&S_3 = P(\mu_2 \leq \Delta\pi^e < \mu_3) = \Phi(\frac{-\Delta\pi^e + \mu_3}{\sigma}) - \Phi(\frac{-\Delta\pi^e + \mu_2}{\sigma}) \\
&S_4 = P(\mu_3 \leq \Delta\pi^e < \mu_4) = \Phi(\frac{-\Delta\pi^e + \mu_4}{\sigma}) - \Phi(\frac{-\Delta\pi^e + \mu_3}{\sigma}) \\
&S_5 = P(\mu_4 \leq \Delta\pi^e) = 1 - \Phi(\frac{-\Delta\pi^e + \mu_4}{\sigma})
\end{aligned} \tag{20}$$

This basic model can, along Kaiser and Spitz (2002), easily be extended to one which can be applied to repeatedly conducted surveys. They assume that the expected variable (Kaiser and Spitz (2002) seek to quantify quarterly revenues of firms by survey data) depends on a constant term β and a disturbance term ϵ. Let I_{jt} be a dummy variable for the participation in the survey of individual j at time $t \in 1...T$, then the equation

[6] For the sake of clarity the notation was simplified in the equations below: t and j subscripts were omitted.

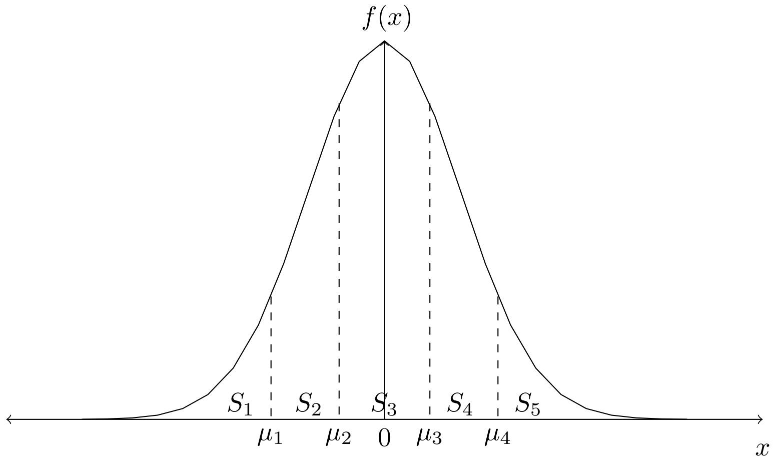

Figure 1: Threshold model with a standard normal distribution

for inflation within the ordered probit model can be specified as $\pi_j t = \sum_{t=1}^{T} \beta_t I_{jt} + \epsilon_{jt}$. Along Kaiser and Spitz (2002), the threshold model, adapted for the here envisaged purpose, is:

$$\pi_{jt}^* = \begin{cases} S_1 & if & & \pi_{jt} = \sum_{t=1}^{T} \beta_t I_{jt} + \epsilon_{jt} & < \mu_1 \\ S_2 & if & \mu_1 \leq & \pi_{jt} = \sum_{t=1}^{T} \beta_t I_{jt} + \epsilon_{jt} & < \mu_2 \\ S_3 & if & \mu_2 \leq & \pi_{jt} = \sum_{t=1}^{T} \beta_t I_{jt} + \epsilon_{jt} & < \mu_3 \\ S_4 & if & \mu_3 \leq & \pi_{jt} = \sum_{t=1}^{T} \beta_t I_{jt} + \epsilon_{jt} & < \mu_4 \\ S_5 & if & & \pi_{jt} = \sum_{t=1}^{T} \beta_t I_{jt} + \epsilon_{jt} & \geq \mu_4 \end{cases} \tag{21}$$

Along the assumptions by Kaiser and Spitz (2002), β_t would then correspond to the expected change in inflation (or, as will be seen later, if applied to question Q5, to the perceived inflation rate). The basic formulation of the threshold model, which only incorporates dummy variables I_{jt} that signal participation, would correspond to the quantification method of Carlson-Parkin with constant threshold values. To however include individual- and time-specific characteristics in the specification would allow for individual- and time-specific threshold values. This suggests to include the various socio-demographic information contained in the ECAMME data set as explanatory variables in the ordered probit estimation.

Kaiser and Spitz (2002) further suggest to interact the time dummy variables with, in their context, firm level dummies to gain firm specific expected revenue growth rates for each quarter. With an average of 1788 individuals per period (after cleaning the data) and without any panel structure, such a model is for obvious reasons not estimable. To overcome this problem this paper extends the approach suggested by Kaiser and Spitz (2002): Using a self-organizing Kohonen map, a pseudo panel is constructed, which allows to follow the inflation perceptions/expectations of different groups over the whole time period available in the dataset. Then the approach described above is implemented in the context of an ordinal MCMC Hierarchical Bayesian Model with a probit link function which allows to estimate the parameter estimate β and thus the expected change of inflation on a cohort level. The approach is outlined in the next two sections.

5. Construction of a pseudo panel

The panel-structure problem is addressed by forming a synthetic or pseudo panel. The first author who used this technique to overcome a lack of panel structure was Deaton (1985). He uses variables that are supposed not to change over time such as sex and birth cohorts to group the survey population. This paper however follows a neural network technique to construct pseudo panels introduced by Gardes et al. (1996) as it exhibits three crucial advantages:

1. It allows for an inclusion of more comprehensive and precise socio-demographic information when building the pseudo panel. Deatons technique applied to the ECAMME dataset with variables as sex (2 categories), region (22 categories) and birth cohorts (3 categories; the survey participants are for example manually grouped into three cohorts: a) under 30 years, b) between 30 and 55 years, c) over 55 years) would result into 132 cohorts. In the neural network approach the number of cohorts can bee freely chosen with respect to the overall survey population. An inclusion of much more socio-demographic information becomes possible. The information content of continuous variables as for example income (such a variable unfortunately is absent in the dataset of ECAMME after 2003) does not have to be artificially reduced by grouping them into different classes, as it would be necessary when using Deaton's approach. They can be used directly in the construction of the pseudo panel.
2. A pseudo panel constructed by the neural network technique is better balanced. The ECAMME data set available exhibits huge imbalances as far as in example the variables sex and region are concerned. The survey is mainly addressed to the head of the household. As a consequence, in around 60% of all cases the gender variable (sex) is male. Similarly, most of the interviewed participants which are randomly chosen from the official French telephone register come from the metropolitan area of Paris (region11), reflecting the fact that around 20% of the French population are living in one of the eight departments of Paris and its surroundings. Interacting these variables into cohorts as Deaton suggests would result in unbalanced cohort sizes, eventually leading to heteroscedasticity in the estimation. This is an issue which does not exist when using the neural network approach by Gardes et al. (1996).
3. Cottrell and Gaubert (2007) show that constructing cohorts via neural networks results in a lower within cohort and a higher between cohort variance than when using the technique by Deaton. This is crucial feature when constructing a pseudo panel.

In this paper, along the work by Gardes et al. (1996) and, Cottrell and Gaubert (2007), self-organizing Kohonen maps are applied to group the participants in the repeated cross

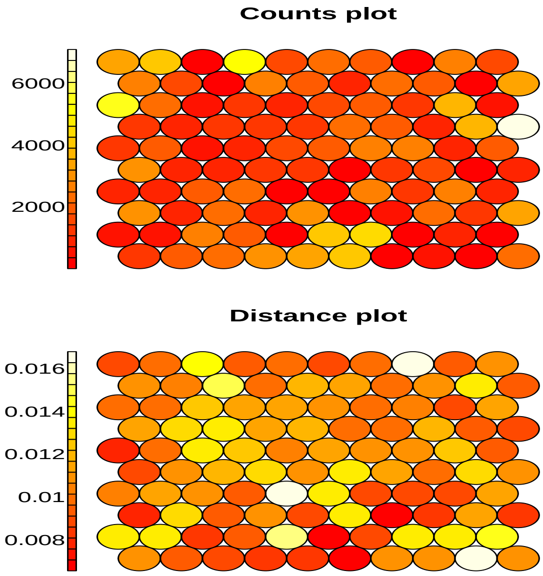

Figure 2: Plot of the self organizing Kohonen map

sections into synthetic cohorts: Socio-demographic variables describing participants in the dataset as sex, birthyear (birthyr), region, citysize, education (educ), childs (number of children), fracwork (the percentage of people in a household who have a job), revquart (the income quartile of the household), workregime (the regime of the employment), finan (the perceived financial situation of the household), spouse, nbpers (number of persons living in the houshold) and occup (occupation) are presented to the neural network in overall hundred runs. The Kohonen map was constructed on a 10×10 hexagonal plane over the whole dataset. Then after it was checked which of the cohorts are available in all time periods. The number of cohorts which are available over the whole time period of the dataset corresponds to 59. To achieve a clean separation of cohorts only the four individuals with the lowest unit to cell distances within each cohort and time period were used. Figure 2 displays the counts for each cell (cohort) over the whole data set and the average distance measures for each cell (cohort). For the construction of the Kohonen map the “kohonen” R-package was used (Wehrens and Buydens, 2007).

6. Estimating cohort-level inflation with ordinal HB-MCMC

Interacting time and cohort dummy variables within an ordered probit estimation in order to gain group/cohort-level estimates of the expected change in inflation rate as suggested by Kaiser and Spitz (2002) is, as already mentioned above, not a good idea in the context of this work. It results in to singularity issues and over-specification. The here chosen approach is to split up the estimation into a fixed effect part, containing several socio-economic variables, and a random effects part which allows to estimate cohort-specific slopes for the time dummy variable corresponding to a cohort-specific estimate of the expected change of inflation. In this respect an ordinal Hierarchical (or Mixed) Bayes Markov Chain Monte Carlo method seems most appropriate as it is widely used for example in marketing studies (*conjoint analysis*) to investigate consumer level reactions to certain product characteristics.

Bayesian estimation, what is it about:. In Bayesian estimation methods the parameter estimates are regarded as random variables while in Maximum likelihood estimations they are viewed as fixed maximizers. This means that all possible values for a parameter estimate θ are compared and ranked. To do so, its distribution conditional on the data x, $p(\theta|x)$ has to be known. This distribution is called the *posterior* distribution which is determined by the *likelihood* $p(x|\theta)$ and the *prior* distribution $p(\theta)$ along $p(\theta|x) = p(x|\theta)p(\theta)/p(x)$. The prior $p(\theta)$ is provided by the user. $p(x)$ is regarded as a normalizing constant, the so called evidence. The estimate of the parameter θ could then be computed by solving the integral $\hat{\theta} = E_{\theta|x}[\theta] = \int \theta p(\theta|x_1, x_2, \dots x_n)d\theta$. At this point however the normalization constant or evidence is missing. It could be calculated by solving the integral $p(x) = \int p(x|\theta)p(\theta)d\theta$. This integral is however only solvable analytically if the *likelihood* and the *prior* form a conjugate pair. If this is not the case, Markov Chain Monte Carlo (MCMC) methods can be used. With MCMC the posterior distribution is evaluated point by point until a constant is reached. In this context the *Metropolis-Hastings algorithm* is normally applied. It randomly draws samples from a proposed distribution. Iteratively points leading to a higher probability are accepted while points leading to a equal or lower probability are rejected. Thereby regions of high probability are iteratively explored. Out of the accepted sample, the parameter can then be estimated by integrating over the posterior distribution (Orbanz, 2013).

The Hierarchical Bayes Markov Chain Monte Carlo method provides the following advantages for there envisioned task:

1. It allows for heterogeneity without having to estimate a model cohort by cohort.
2. It can handle relatively small amounts of cohort-level data and still gives good estimates as it borrows information from the whole population to gain individual-level or here cohort-level estimates.
3. One can run very complex estimations. The inclusion of explanatory variables is only limited by hardware restrictions.

6.1. Specification

The estimation done in this paper were conducted with R-package “MCMCglmm” by Hadfield (2010). The setup was as follows:

$$Y_{tj} = X_{tj}\beta + Z_{tj}b_{tj} + e_{tj} \tag{22}$$

where X is the design matrix for the fixed effects with parameters β, and Z is the design matrix for the random part with parameters b. j denotes the j^{th} cohort. The following specification was used for the fixed effects:

$$
\begin{aligned}
&X_{tj}\beta = \\
&\beta_1 SEX_{tj} + \beta_2 EDUC_{ti} + \beta_3 REGION_{tj} + \beta_4 AGE_{tj} + \beta_5 CHILDS_{tj} \\
&+ \beta_6 REVQUARTtj + \beta_7 REGION_{tj} + \beta_8 CITYSIZEtj + \beta_9 WORKREGIMEtj \\
&+ \beta_{10} FRACWORKtj + \beta_{11}\text{ECON\_PER}tj + \beta_{12}\text{ECON\_EXP}tj + \beta_{13} FINANtj
\end{aligned}
$$

- SEX ... is a categorical variable in the survey denoting the sex of the interviewed head of the household
- $EDUC$ ... is a categorical variable in the survey regarding the education of the respondent
- $REGION$ ...is a categorical variable in the survey regarding the region of residence of the respondent
- AGE ... the age of the respondent
- $CHILDS$ ... number of children below 14 years living in the same household as the respondent
- $REVQUART$ ... is a categorical variable describing the income quartile to which a household belongs
- $CITYSIZE$ ... categorical variable describing the size of the city in which the household has its residence
- $FRACWORK$ ... a numeric variable between 0 and 1 describing the percentage of people within the household with a job
- $ECON_PER/ECON_EXP$ ... categorical variables describing the perceptions/expectations of households with regard to the French economy
- $FINAN$ ... categorical variables describing the financial situation as perceived by the household itself

The random effect part is specified as follows:

$$Z_{tj}b_tj = b_{tj}ID_{tj} * WAVE_{tj} \tag{23}$$

- $WAVE$ ... dummy variable for the respective survey wave. In this part the WAVE-specific random intercept is estimated.
- ID ... is the variable denoting the cohort ID.
- $ID_{tj} * WAVE_{tj}$ ... This part allows the cohort-level random intercepts to be estimated.

Along Berk (1999) the expected changes in inflation are calculated in two steps: As outlined in Section 4.2, the expected inflation in the Carlson-Parkin approach is calculated on the basis of a perceived inflation rate. For the sake of simplicity in the literature the perceived inflation rate is normally assumed to be the latest published inflation rate. This assumption is however debatable. As the ECAMME contains a question which asks the respondents for their inflation perceptions (see Section 2, question Q5), one can estimate a perceived inflation along the same method on which one can then base the estimated expected inflation rate. The model described above is thus estimated for two dependent variables:

1. the perceived inflation given by the 5-category variable I^p_{tj} (see Section 2, Q5) — the perception estimate (random intercept) is further on denoted as b_{1tj};
2. the expected inflation rate given by the 5-category variable I^e_{tj} (see Section 2, Q6) — the expectation estimate (random intercept) is further on denoted as b_{2tj};

6.2. Settings and Diagnostics

The priors for the fixed effect structure (B), the variance structure of the residuals (R) and the variance structure of the random part (G) are specified as follows to run the ordinal Hierarchical Bayesian estimation,

$$B \sim \mathrm{N}\left(0, diag(dim(X)) * 1e10\right)$$
$$R \sim \mathrm{W}^{-1}\left(V = 1, nu = 1\right)$$
$$G \sim \mathrm{W}^{-1}\left(V = diag(N), nu = N + 2\right)$$

where N denotes the number of cohorts and $dim(X)$ denotes the dimension of the fixed effects model matrix. W^{-1} denotes the Inverse-Wishart distribution. The R structure priors are set along (Hadfield, 2010, Table 1) for ordinal regressions.

	post. mean	low. 95 perc. conf. lim.	up. 95 perc. conf. lim.	effect. sample	pMCMC	signif.
age	0.002	-0.000	0.005	5038.271	0.103	
econ_per1	3.594	2.890	4.253	3595.727	0.000	***
econ_per2	3.167	2.627	3.755	3573.543	0.000	***
econ_per3	3.229	2.692	3.805	3482.898	0.000	***
econ_per4	3.688	3.123	4.243	3433.360	0.000	***
econ_per5	4.364	3.806	4.918	3359.485	0.000	***
econ_exp.1	-0.323	-0.527	-0.117	4535.841	0.003	**
econ_exp.2	0.205	0.028	0.371	4800.000	0.023	*
econ_exp.3	-0.014	-0.122	0.094	4800.000	0.811	
econ_exp.4	-0.035	-0.091	0.025	4800.000	0.227	
region21	-0.021	-0.269	0.226	5465.550	0.870	
region22	0.133	-0.106	0.386	4498.877	0.278	
region23	0.154	-0.072	0.388	5069.825	0.190	
region24	0.148	-0.070	0.358	4503.697	0.184	
region25	0.065	-0.180	0.313	4800.000	0.609	
region26	0.050	-0.182	0.285	6130.088	0.685	
region31	0.101	-0.110	0.310	4800.000	0.352	
region41	0.221	0.009	0.450	4800.000	0.049	*
region42	0.098	-0.141	0.318	4800.000	0.412	
region43	0.042	-0.233	0.287	4800.000	0.735	
region52	-0.015	-0.221	0.206	4597.017	0.887	
region53	0.020	-0.194	0.217	4800.000	0.859	
region54	0.007	-0.220	0.256	4564.842	0.942	
region72	0.150	-0.060	0.365	4800.000	0.161	
region73	0.074	-0.142	0.279	4800.000	0.501	
region74	-0.064	-0.338	0.238	4800.000	0.666	
region82	0.240	0.041	0.439	4800.000	0.022	*
region83	0.341	0.113	0.600	4800.000	0.006	**
region91	0.077	-0.142	0.322	4800.000	0.513	
region93	0.319	0.107	0.517	4800.000	0.003	**
region94	0.528	0.018	1.047	4800.000	0.049	*
citysize.1	-0.062	-0.196	0.077	4545.323	0.380	
citysize.2	-0.017	-0.181	0.126	4800.000	0.829	
citysize.3	0.125	-0.045	0.273	4800.000	0.122	
citysize.4	0.061	-0.086	0.200	4800.000	0.393	
citysize.5	-0.071	-0.206	0.073	4588.974	0.322	
citysize.6	0.038	-0.120	0.179	4800.000	0.612	
citysize.7	-0.114	-0.262	0.053	4800.000	0.168	
citysize.8	0.126	-0.038	0.291	4800.000	0.142	
educ.1	-0.227	-0.311	-0.143	4800.000	0.000	***
educ.2	-0.116	-0.184	-0.051	5031.206	0.002	**
educ.3	-0.148	-0.205	-0.089	5045.272	0.000	***
childs	0.029	-0.010	0.065	4800.000	0.139	
fracwork	0.032	-0.092	0.167	4800.000	0.620	
sex2	0.297	0.221	0.372	4800.000	0.000	***
revquart.1	-0.086	-0.168	-0.002	4800.000	0.043	*
revquart.2	-0.029	-0.095	0.039	5078.338	0.381	
revquart.3	-0.046	-0.100	0.014	4800.000	0.120	
workregime1	0.098	-0.058	0.243	4632.493	0.214	
workregime2	0.189	-0.092	0.489	4800.000	0.203	
workregime9	-0.124	-0.273	0.040	4800.000	0.122	
finan2	0.227	-0.213	0.674	5532.799	0.334	
finan3	0.469	0.028	0.926	5524.299	0.040	*
finan4	0.593	0.109	1.019	5509.132	0.008	**
finan5	0.878	0.353	1.411	6949.582	0.002	**
cutpoint.1	2.266	2.155	2.373	133.383		
cutpoint.2	3.805	3.688	3.912	131.167		
cutpoint.3	5.083	4.964	5.193	129.275		

Table 3: Perception estimation - Fixed effects [Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1]

	post. mean	low. 95 perc. conf. lim.	up. 95 perc. conf. lim.	effect. sample	pMCMC	signif.
age	0.003	0.001	0.006	4800.000	0.025	*
econ_per1	3.535	2.855	4.211	4800.000	0.000	***
econ_per2	3.640	3.073	4.209	4800.000	0.000	***
econ_per3	3.557	2.994	4.103	3068.690	0.000	***
econ_per4	3.618	3.049	4.163	3009.613	0.000	***
econ_per5	3.672	3.102	4.215	3011.347	0.000	***
econ_exp.1	-1.257	-1.456	-1.056	4800.000	0.000	***
econ_exp.2	0.037	-0.132	0.208	4800.000	0.685	
econ_exp.3	0.056	-0.051	0.165	4552.892	0.315	
econ_exp.4	-0.048	-0.102	0.009	4338.935	0.097	.
region21	0.125	-0.128	0.369	4800.000	0.305	
region22	-0.052	-0.289	0.183	4048.983	0.677	
region23	0.122	-0.108	0.340	4800.000	0.305	
region24	0.142	-0.068	0.357	4800.000	0.175	
region25	0.151	-0.104	0.385	4800.000	0.225	
region26	0.023	-0.192	0.263	4034.229	0.836	
region31	-0.101	-0.295	0.114	4246.428	0.343	
region41	-0.207	-0.417	0.012	5002.027	0.063	.
region42	-0.258	-0.477	-0.026	5631.955	0.027	*
region43	-0.011	-0.282	0.241	4800.000	0.927	
region52	0.121	-0.078	0.324	4272.908	0.247	
region53	0.246	0.043	0.452	4800.000	0.015	*
region54	0.060	-0.166	0.296	4800.000	0.622	
region72	-0.022	-0.224	0.187	4800.000	0.850	
region73	0.128	-0.086	0.335	4186.523	0.230	
region74	0.298	0.022	0.593	4800.000	0.039	*
region82	0.129	-0.058	0.316	4176.320	0.189	
region83	0.206	-0.033	0.438	4800.000	0.095	.
region91	0.088	-0.133	0.319	5067.657	0.446	
region93	0.127	-0.069	0.331	4105.664	0.213	
region94	0.227	-0.238	0.724	4800.000	0.372	
citysize.1	-0.087	-0.221	0.049	4800.000	0.212	
citysize.2	-0.027	-0.178	0.119	4800.000	0.720	
citysize.3	0.178	0.025	0.329	4800.000	0.021	*
citysize.4	-0.035	-0.178	0.100	4800.000	0.621	
citysize.5	-0.028	-0.160	0.100	4888.152	0.679	
citysize.6	-0.033	-0.174	0.125	4800.000	0.663	
citysize.7	0.070	-0.072	0.232	4800.000	0.386	
citysize.8	-0.054	-0.217	0.101	4800.000	0.511	
educ.1	0.099	0.020	0.188	4259.146	0.020	*
educ.2	-0.058	-0.129	0.005	4800.000	0.092	.
educ.3	0.049	-0.004	0.110	4800.000	0.086	.
childs	-0.018	-0.053	0.019	4800.000	0.334	
fracwork	0.001	-0.126	0.122	4800.000	0.977	
sex2	-0.038	-0.110	0.037	4799.234	0.320	
revquart.1	-0.080	-0.160	0.001	4800.000	0.054	.
revquart.2	-0.045	-0.114	0.018	4800.000	0.172	
revquart.3	-0.014	-0.070	0.040	4384.701	0.641	
workregime1	0.156	0.016	0.296	4130.997	0.027	*
workregime2	-0.062	-0.338	0.215	4800.000	0.659	
workregime9	-0.052	-0.195	0.083	4800.000	0.462	
finan2	-0.355	-0.844	0.104	4800.000	0.140	
finan3	-0.234	-0.730	0.224	4800.000	0.334	
finan4	-0.228	-0.718	0.236	4800.000	0.353	
finan5	0.092	-0.452	0.612	4800.000	0.737	
cutpoint.1	2.675	2.605	2.755	232.070		
cutpoint.2	3.430	3.353	3.506	221.452		
cutpoint.3	6.051	5.968	6.137	235.715		

Table 4: Expectation estimation - Fixed effects [Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1]

The estimations were run with 300,000 iterations each and with the thinning and burning parameters set to 50 and 60,000 respectively. The estimation of both models on a computer with a i7-2640M processor and 8GB of memory took around 36 hours.

The crucial results of the estimation process are the posterior distributions of the random intercepts. Integrating over the posterior distributions of the random intercepts supplies the "BLUPs" (linear unbiased predictors) or "conditional modes" and thus the individual level parameter estimates. The two parameters estimates per period and cohort can then be used to calculate the perceived and expected inflation rate respectively for each cohort in each period. The calculation of the quantified measure for the perceived and the expected inflation will be discussed in the next section.

Convergence. In order to check convergence of the estimated chains the Heidelberger and Welch's convergence diagnostic (Heidelberger and Welch, 1981) is applied to both estimations. The diagnostic consists of two parts: first a Cramer-von-Mises statistic is applied to check the null hypothesis that the sample values are drawn from a stationary distribution. Therefore this test is at first conducted on the whole sample and then successively the first 10%, 20% and so on of the chain are cast aside until the H_0 is accepted. If the H_0 cannot be accepted before 50% are cast aside, the diagnostic fails. In this case one would had to rerun the estimation with more iterations in order to achieve stationary posterior distributions. In a second part of the diagnostic the 95% confidence intervals for the means of the fractions of the posterior distributions for which the H_0 of stationarity was accepted are computed. In the case the ratio between the half width or radius of this confidence interval and the estimated mean is lower than a target value the sample size is regarded as being too small to estimate the mean with sufficient precision. The target value for the ratio of halfwidth to the sample mean was set to be 0.3 and the p-value was set to 0.05.

		stationarity test (pct. passed)	start iteration	combined p-value (Stouffer)	halfwidth test (pct. passed)	avg. mean	avg. halfwidth
Perceptions	fixed	0.982	98.778	0.036	0.964	0.417	0.004
	random	0.993	98.778	0.989	0.933	0.011	0.018
Expectations	fixed	1	62.091	0.088	0.964	0.319	0.004
	random	0.993	62.091	1	0.922	0.009	0.018

Table 5: Heidelberger/Welch Stationarity Test and Halfwidth Test

The results of the Heidelberger and Welch's convergence diagnostic are displayed in Table 5: The convergence diagnostic states that the estimated posterior distributions are stationary in more than 99% of all cases. The Halfwidth Test passed in more than 92% in all cases, indicating that the models were run with a sufficient amount of iterations and that the means of the posterior distributions have been estimated with adequate accuracy for nearly all parameter estimates. Only a negligible proportion of posterior distributions thus failed to converge.

7. Estimation results and computation of the perceived/expected inflation rate

7.1. Discussion of the raw perception/expectation estimates

Figure 3 compares the actual inflation rate (first graph) with the perception estimate (b_{1tj}, second graph) and the expectation estimate (b_{2tj}, third graph): The second and the third graph display monthly averages over all cohorts for the random intercept estimates of the two (perception/expectations) hierarchical MCMC ordered probit models presented in the previous section.

The synchronous behavior of the perception estimate with the actual inflation rate, even though both measures are on a different scale, is apparent on the first glance.

This is verified by Figure 4: Computing the correlation between the actual inflation rate and the perception estimate with different lags not only shows that respondents have a very good intuition for past price changes but also perceive these price changes promptly after they occur: The highest correlation between the two time series is at a lag of zero

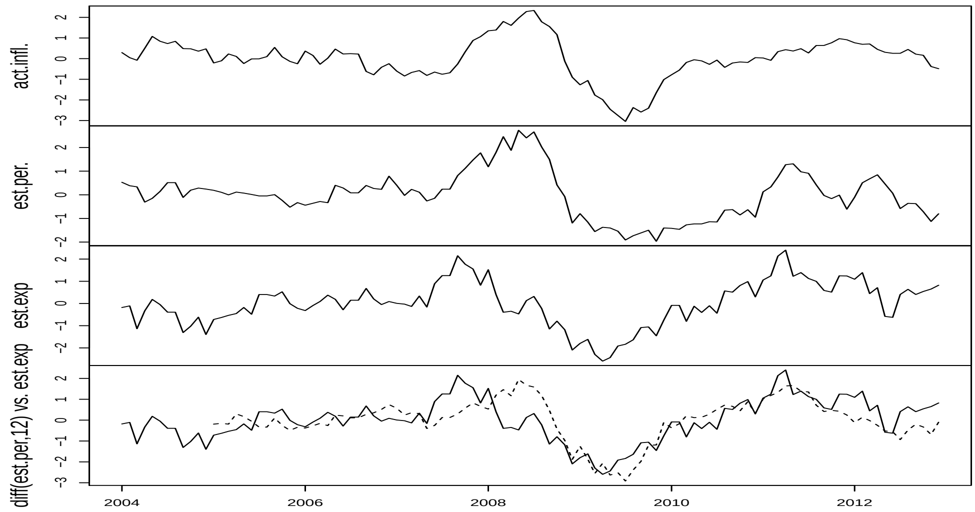

Figure 3: Raw estimates of perceptions and expectations; differentiated perceptions vs. expectations - Jan 2004 to Dec 2012

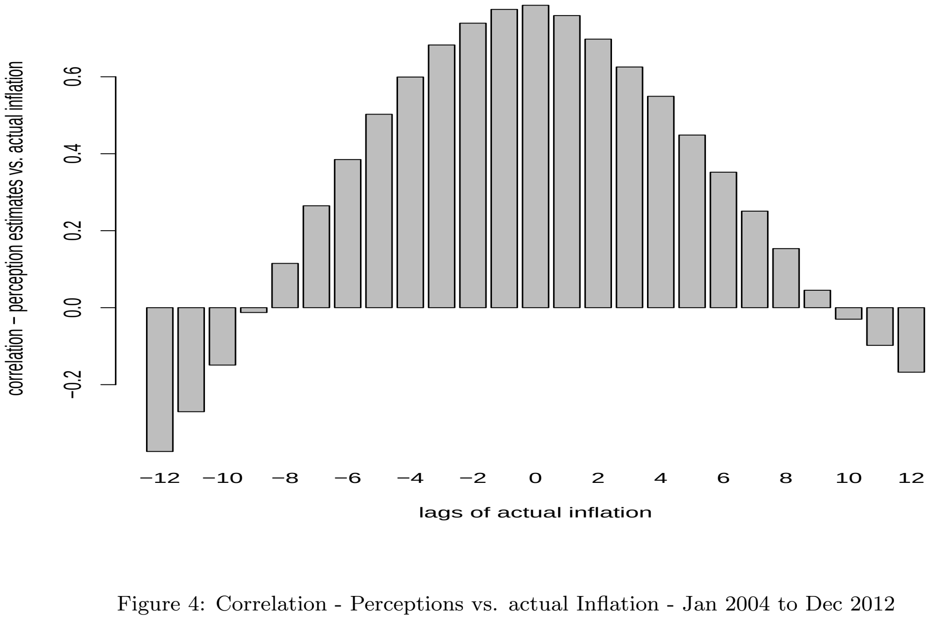

Figure 4: Correlation - Perceptions vs. actual Inflation - Jan 2004 to Dec 2012

and one (this means where actual inflation rate from $t-1$ is compared with the perception estimate from t). It has to be again underlined that the perception estimate solely bases on the mean of the posterior distributions of the cohort/period-level random intercepts estimated with the model with Q5 (see Section 2) as the dependent variable as outlined in the last section. The inflation rate itself is not part of this estimation in any form.

Comparing the graph of the actual inflation rate (graph 1) and the expectation estimate (graph 3) in Figure 3, such a striking synchrony cannot be observed. Rather the expectation estimate, which corresponds to the mean of posterior distributions of cohort/period-level random intercepts of estimation 2 with variable Q6 (see Section 2) as dependent variable, seems to precede the actual inflation by a view months. This assertion is however misleading. To see that, one has to take a closer look at the wording of the questions Q5, about the perceived past inflation, and question Q6, about the expected future inflation (see Section 2): The backward looking question asks for the "price change" in the last twelve months. This corresponds to the price change of the consumer basket as surveyed by the national statistics office in terms of percentages between today and the same period (month) one year before. The forward looking question Q6, in contrast, asks for the evolution of prices in the next twelve months compared with the last twelve months. Response options here are for example: the prices will rise with a *higher* rate, with the s*ame* rate, with a *smaller* rate and so on. One could thus say that the backward looking question (Q5) asks for the derivative of consumer price of the first order, while the forward looking question (Q6) asks for the second order derivative of prices (the expected change of price changes).

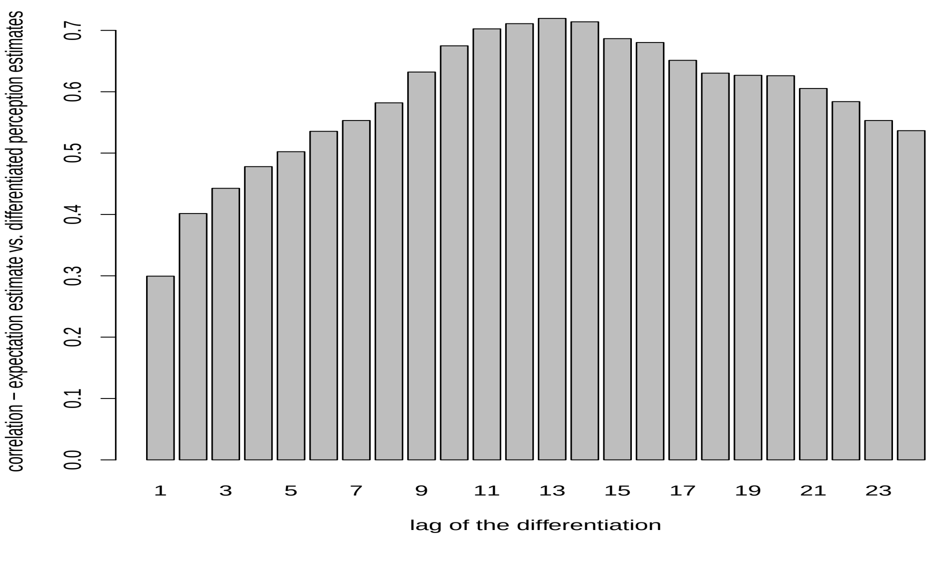

Figure 5: Correlation - Expectations vs. differentiated Perceptions - Jan 2004 to Dec 2012

This difference in wording is also reflected in the estimates for perceptions and expectations computed by the method outlined in the last section. The fourth graph in Figure 3 displays the expectation estimate and the perception estimate, where the latter was transformed by twelve month differences as suggested by the wording of the question to get the two series on the same order: by this procedure we get a time series reflecting the perceived changes of prices changes (inflation) in the last twelve months. Figure 5 displays the correlation between the expectation estimate on the one hand and the perception estimate with different difference lags on the other hand. According to the wording of question Q6 the correlation is the highest when the perception estimate is transformed by a twelve months difference.

Corresponding to these assertions, the technique to compute estimates for the perceived and the expected inflation rates which are comparable to the actual inflation rate and thus practical for the application of the non-parametric test for herd behavior described in section 3 is outlined in the next subsection.

7.2. Deriving the expected inflation rate

As discussed in the last subsection the perception estimate is closely correlated to the actual inflation rate: Respondents have a good intuition of price changes in terms of tendencies even if they occurred quite recently. The actual inflation rate and the perception estimate as well as the expectation estimate are however on different scales. This problem is solved by standardizing all measures to $\mu = 0$ and $\sigma = 1$. This procedure was conducted for each cohort time series separately.

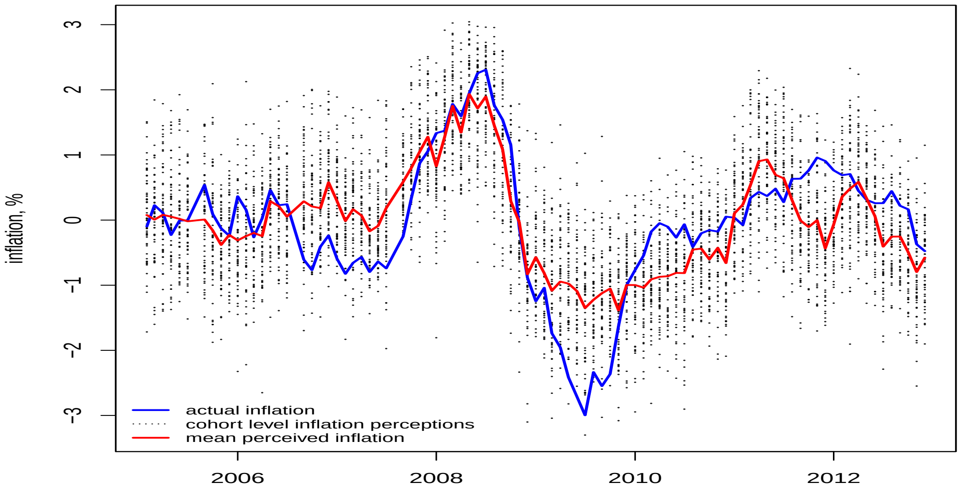

Figure 6: Cohort-level inflation perceptions and actual inflation - Jan 2004 to Dec 2012

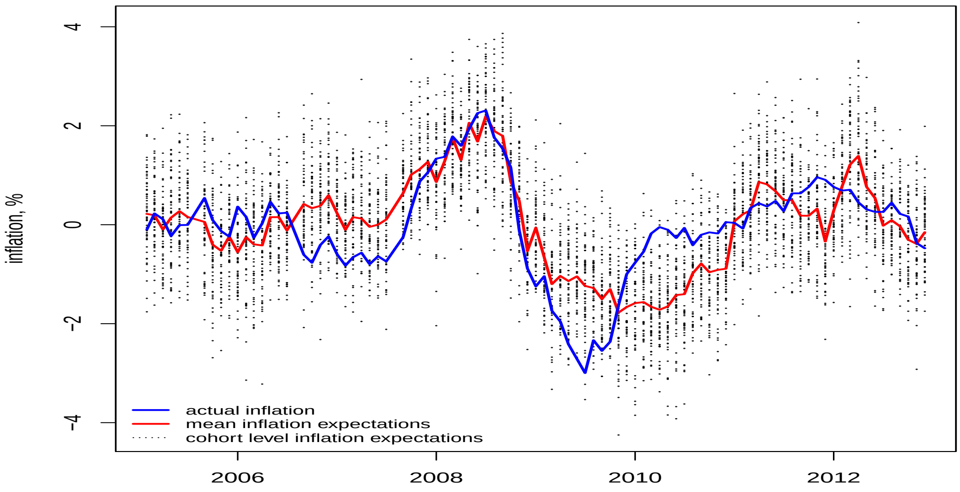

Figure 7: Cohort-level inflation expectations and actual inflation - Jan 2004 to Dec 2012

In the case of the expectation estimate, the time series for each cohort was integrated with a twelve months difference before the procedure, to get the different variables (perception and expectation estimate, actual inflation) to the same order and re-transformed by taking the twelve months difference afterwards. Figure 6 gives an impression about the distribution of cohort-level inflation perceptions for each cohort in the data set and over all periods.

The perceived inflation rate computed by normalization is denoted as π^p_{tj} . Along the wording of question Q6 and what was shown in the last subsection it is assumed that the expectation estimate reflects the change of the perceived inflation between t and $t + 12$. The expected inflation rate π^e_{tj} can thus be calculated by:

$$\pi^e_{tj} = \pi^p_{tj} + b^2_{1tj} \tag{24}$$

b_{1tj} and b_{2tj} here denote the respective wave/cohort-level intercepts from estimation 1 (perceptions) and 2 (expectations; see Section 6) respectively. Figure 7 displays the distribution of cohort-level inflation expectations in comparison with the actual inflation rate and the mean of the expected inflation rate.

Figures D.9. and D.10 display charts of the perceived and the expected inflation rates for each cohort against the actual inflation rate.

The procedure outlined above pursues the goal to make the data applicable to the herding test statistics. To obtain quantitative values of the perceived and expected inflation in different context which correspond to the scale of the actual inflation rate one could simply scale the cohort-level perception estimates in each period to the latest published actual inflation rate. The expected inflation rate can then be computed, analogously to above, by adding the expectation estimate to the perceived inflation rate in t to get the expected inflation rate for $t + 12$. This again has to be done on a period/cohort-level.

Another idea would be, to use the quantitative questions with regard to inflation perceptions/expectations and to compare them with the qualitative answers in order to calculate thresholds which can then be used to estimate the model. This would however mean to fix the thresholds for cohorts over time, since the values stated by most individuals are often not really realistic (this was already discussed above). In the ordered probit framework used here the thresholds are estimated within the model.

8. Application of the test statistics

With the estimated cohort level inflation expectations at hand, computed along the method described above, the test statistics outlined in Section 3 can be applied.

ECAMME asks the respondents for the inflation rate twelve months ahead. Therefore, according to the wording of question Q6 (see Section 2) the inflation rate from twelve months ahead, after the date of the interview (interview *wave*), is used as the actual or target inflation rate which participants are asked to forecast and then, along the test statistic, compared with the current individual-level inflation expectations and consensus expectations.

The test statistic is applied quarterly: For the first month of each quarter the consensus forecast for one cohort is the by cohort size (the cohort size normally is four, but if the cohort after clustering had a lower count then four this number can also be lower) weighted arithmetic mean of forecasts from all other cohorts in the respective month. In the second month of each quarter the consensus is the by cohort size weighted arithmetic mean of forecasts from all other cohorts in the first month. In the third month of the quarter the consensus is the by cohort size weighted arithmetic mean of forecasts from all other cohorts in the first month and the second month of the respective quarter. The monthly results for the test statistic are computed and averaged over the quarter.

Date	$Pr(F_t > E_t\|z_t^+)$	$Pr(F_t < E_t\|z_t^-)$	S	S (lower bound)	S (upper bound)
2005 Q2	1.000	0.850	0.925	0.886	0.964
2005 Q3	0.848	0.925	0.868	0.842	0.894
2005 Q4	0.976	0.934	0.955	0.915	0.994
2006 Q1	0.571	1.000	0.786	0.746	0.825
2006 Q2	0.608	1.000	0.804	0.765	0.843
2006 Q3	0.924	0.574	0.730	0.704	0.756
2006 Q4	1.000	0.373	0.687	0.648	0.726
2007 Q1	1.000	0.274	0.637	0.598	0.676
2007 Q2	1.000	0.342	0.671	0.632	0.710
2007 Q3	1.000	0.364	0.773	0.747	0.799
2007 Q4	0.875	0.931	0.903	0.864	0.943
2008 Q1	0.770	1.000	0.885	0.846	0.924
2008 Q2	0.716	0.988	0.852	0.813	0.892
2008 Q3	0.893	0.875	0.884	0.845	0.923
2008 Q4	1.000	0.398	0.699	0.660	0.738
2009 Q1	1.000	0.325	0.663	0.624	0.702
2009 Q2	1.000	0.142	0.571	0.532	0.610
2009 Q3	1.000	0.131	0.566	0.527	0.605
2009 Q4	0.860	0.675	0.767	0.729	0.806
2010 Q1	0.232	1.000	0.616	0.577	0.655
2010 Q2	0.095	1.000	0.548	0.509	0.587
2010 Q3	0.214	1.000	0.607	0.568	0.646
2010 Q4	0.283	1.000	0.641	0.602	0.680
2011 Q1	0.615	1.000	0.808	0.769	0.847
2011 Q2	0.958	0.929	0.943	0.904	0.983
2011 Q3	0.985	0.828	0.907	0.868	0.946
2011 Q4	0.394	1.000	0.697	0.658	0.736
2012 Q1	0.619	1.000	0.809	0.770	0.848
2012 Q2	1.000	0.494	0.747	0.708	0.786
2012 Q3	0.948	0.907	0.927	0.888	0.967
2012 Q4	0.969	0.848	0.908	0.869	0.947

Table 6: Results - test statistics

9. Conclusion

Table 6 and Figure 8 represent the results of the herding test: As outlined in Section 3, a S value of below 0.5 would indicate herding behavior. A value S of above 0.5 on the other hand indicates anti-herding behavior.

The hypothesis that there is herding behavior within consumer expectations towards the consensus can be rejected in all periods. More over most periods exhibit a value $S > 0.5$ which indicates anti-herding. Anti-herding corresponds, as already mentioned before, to an overweighting of individual information compared to public information (the consensus).

This result is supported by Rülke and Tillmann (2011) who find that members of the Federal Open Market Committee exhibit a strong anti-herding behavior when uttering their inflation forecasts. This phenomenon is especially strong within non-voting members of the FOMC. Rülke and Tillmann (2011) explain their results as some kind of strategic behavior with regard to monetary policy. This interpretation might of course not be directly applied to the case of household expectations, but the findings of the paper at hand might base on similar mechanisms. Being interviewed by a public entity, respondents might use there answers to implicitly transport their opinion about economic policy related to price changes eventually trying to push it into a certain direction.

One objection might be that the results strongly depend on the purchasing power of respondents. This means that a respondent with lower income will classify a price change, be it perceived or expected, of let's say 1.7 % per annum (the mean of inflation rate in the last twenty years in France) differently on a discrete scale with five categories than a respondent with a higher income. Separating individuals into groups by clustering methods as the self-organizing Kohonen map would thus lead to a result in which inflation expectations strongly relate to personal experiences and information, appearing as

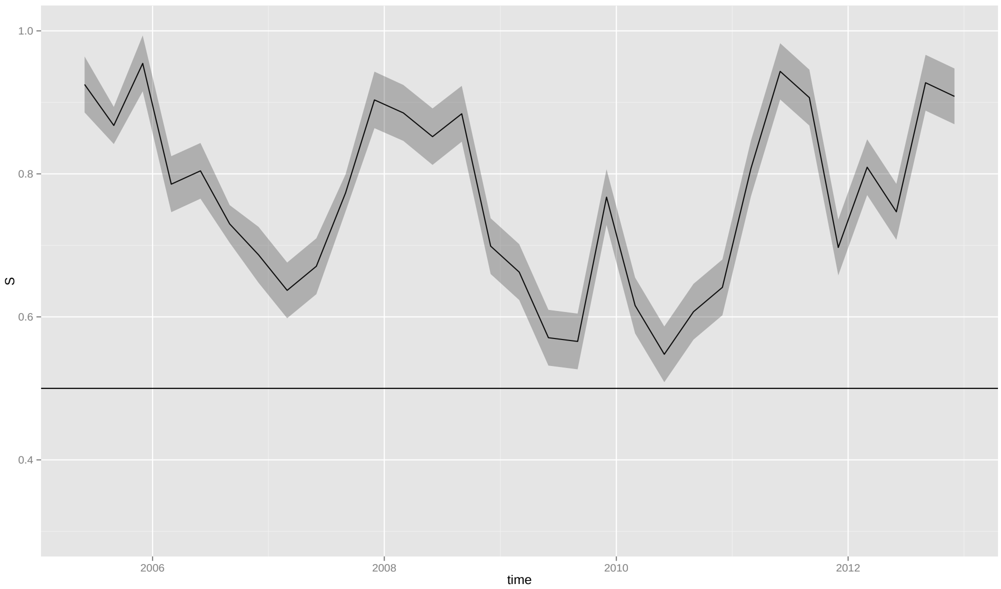

Figure 8: Quarterly Herding Test Statistic - Jan 2004 to Dec 2012 (5% confidence interval)

anti-herding in the results above. This was however controlled for by, first, including the socio-demographic information into the fixed effects to distill the cohort/period-level perceptions and expectations, second, in contrast to other works using pseudo panelization, the data was not aggregated on a cohort level. This means the cohort/period-level expectation and perception estimates are based on the socio-demographic variation within each cohort.

A caveat for the paper at hand and the results presented above lies in the data however. Bernardt et al. (2006) have, as pointed out above, designed the here applied non-parametric test for herding to test the forecasting behavior of professional analysts. This kind of data has the following advantages: first it has a real panel structure over several years, second the exact point of time is known when each analyst published his/her forecast, thus the sequence of forecasts is known. This is unfortunately not the case in the French household survey data (in example unlike to the Michigan consumer survey) where only the month in which the survey took place is known. Therefore a compromise had to be found, namely to apply the test statistic quarterly, using the mean of all prior forecasts within this month as the consensus and comparing it with the cohort-level expectations and the target inflation rate from one year ahead. This implies that a possible adjustment to a consensus, if it takes place, could only be measured once a month and thus three times a quarter. One cannot completely exclude the possibility of different results if this information is available. To further investigate expectation formation with regard to inflation in the context of group or herd behavior, and therefore the timing of expectation formation, it would be desirable to include this information into the data sets of household surveys within the harmonized European household/consumer survey program or to start recording this information when interviews are conducted.

Acknowledgements

I would like to thank François Gardes and Christophe Starzec for their helpful comments, as well as my discussants at the conference of the Société canadienne de science économique 2013 (Quebec Ville, Canada) and at the Journées de Microéconomie Appliqué 2013 (Nice, France).

10. References

Anderson, O., 1952. The business test of the ifo-institute for economic research, munich, and its theoretical model. Revue de l'Institut International de Statistique, 1–17.

Ang, A., Bekaert, G., Wei, M., 2007. Do macro variables, asset markets, or surveys forecast inflation better? Journal of Monetary Economics 54 (4), 1163–1212.

Baghestani, H., 2009. Survey evidence on the muthian rationality of the inflation forecasts of us consumers. Oxford Bulletin of Economics and Statistics 54 (2), 173–186.

Berk, J., 1999. Measuring inflation expectations: a survey data approach. Applied Economics 31 (11), 1467–1480.

Bernardt, D., Campello, M., Kutsoati, E., 2006. Who herds? Journal of Financial Economics 80, 657–675.

Carlson, J., Parkin, M., 1975. Inflation expectations. Economica 42 (166), 123–138.

Cottrell, M., Gaubert, P., 2007. Efficient estimators: the use of neural networks to construct pseudo panels. arXiv preprint math/0701154.

Curto Millet, F., 2006. Finding the optimal method of quantifying inflation expectations on the basis of qualitative survey data. Narodowy Bank Polski.

Deaton, A., 1985. Panel data from time series of cross-sections. Journal of econometrics 30 (1), 109–126.

Forsells, M., Kenny, G., 2002. The rationality of consumers' inflation expectations: survey-based evidence for the euro area. European Central Bank.

Franke, R., 2007. Estimation of a microfounded herding model on german survey expectations.

Friedman, M., 1968. The role of monetary policy. The American Economic Review 58, 1–17.

Gardes, F., Gaubert, P., Rousset, P., 1996. Cellulage de données d?enquêtes de consommation par une méthode neuronale. Rapport de recherche pour le CREDOC.

Gardes, F., Ghabri, S., Madre, J., Pichery, M., 2000. Rationality of price and unemployment expectations: tests on french qualitative microdata. In: Gardes, F., Prat, G. (Eds.), Price expectations in goods and financial markets. Edward Elgar, pp. 131–155.

Gardes, F., Madre, J., 1991. Les anticipations des ménages dans les enquêtes de conjoncture de l'insee. Economie et Prévision 99 (3), 1–29.

Gourieroux, C., Pradel, J., 1986. Direct test of the rational expectation hypothesis. European Economic Review 30 (2), 265–284.

Hadfield, J. D., 2010. Mcmc methods for multi-response generalized linear mixed models: The MCMCglmm R package. Journal of Statistical Software 33 (2), 1–22.
URL http://www.jstatsoft.org/v33/i02/

Heidelberger, P., Welch, P. D., 1981. A spectral method for confidence interval generation and run length control in simulations. Communications of the ACM 24 (4), 233–245.

Howrey, E., 2001. The predictive power of the index of consumer sentiment. Brookings papers on economic activity 2001 (1), 175–216.

Kaiser, U., Spitz, A., 2002. Quantification of qualitative data using ordered probit models. In: Economic surveys and data analysis: CIRET Conference proceedings, Paris 2000. Organization for Economic, p. 325.

Lucas, R., 1976. Econometric policy evaluation: A critique. In: Carnegie-Rochester Conference Series on Public Policy. Vol. 1. pp. 19–46.

Ludvigson, S., 2004. Consumer confidence and consumer spending. The Journal of Economic Perspectives 18 (2), 29–50.

Maag, T., 2009. On the accuracy of the probability method for quantifying beliefs about inflation. ETH, Eidgenössische Technische Hochschule Zürich, KOF Swiss Economic Institute.

Maddala, G., 1991. Survey data on expectations: what have we learnt? Issues in Contemporary Economics 2, 319–344.

Mankiw, N., Reis, R., 2002. Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve*. Quarterly Journal of Economics 117 (4), 1295–1328.

Mehra, Y., 2002. Survey measures of expected inflation: revisiting the issues of predictive content and rationality. Economic Quarterly-Federal Reserve Bank of Richmond 88 (3), 17–36.

Mourougane, A., Roma, M., 2003. Can confidence indicators be useful to predict short term real gdp growth? Applied Economics Letters 10 (8), 519–522.

Muth, J., 1961. Rational expectations and the theory of price movements. Econometrica: Journal of the Econometric Society, 315–335.

Nardo, M., 2003. The quantification of qualitative survey data: a critical assessment. Journal of economic surveys 17 (5), 645–668.

Orbanz, P., 2013. Bayesian estimation. [coursenotes].
URL http://stat.columbia.edu/~porbanz/teaching/slides_ml__bayesian_estimation.pdf

Pesaran, M., 1985. Formation of inflation expectations in british manufacturing industries. The Economic Journal 95 (380), 948–975.

Pesaran, M., 1989. The limits to rational expectations. Cambridge Univ Press.

Pesaran, M., Weale, M., 2006. Survey expectations. Handbook of Economic Forecasting 1, 715–776.

Phillips, A., 1958. The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–1957. Economica 25 (100), 283–299.

Roberts, J., 1997. Is inflation sticky? Journal of Monetary Economics 39 (2), 173–196.

Rülke, J.-C., Tillmann, P., 2011. Do fomc members herd? Economics Letters 113 (2), 176–179.

Sargent, T., Fand, D., Goldfeld, S., 1973. Rational expectations, the real rate of interest, and the natural rate of unemployment. Brookings Papers on Economic Activity 1973 (2), 429–480.

Theil, H., 1952. On the time shape of economic microvariables and the munich business test. Revue de l'Institut International de Statistique, 105–120.

Thomas, L., 1999. Survey measures of expected us inflation. The Journal of Economic Perspectives 13 (4), 125–144.

Vuchelen, J., 2004. Consumer sentiment and macroeconomic forecasts. Journal of economic psychology 25 (4), 493–506.

Vuchelen, J., Praet, P., 1984. The contribution of ec consumer surveys in forecasting consumer expenditures: An econometric analysis for four major countries. Journal of economic psychology 5 (2), 101–124.

Wehrens, R., Buydens, L., 2007. Self- and super-organising maps in r: the kohonen package. J. Stat. Softw. 21 (5).
URL http://www.jstatsoft.org/v21/i05

10.1. Data References

Enquête de conjoncture auprès des ménages (ECAMME) - 2004 - (2004) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2005 - (2005) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2006 - (2006) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2007 - (2007) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2008 - (2008) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2009 - (2009) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2010 - (2010) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2011 - (2011) [fichier électronique], INSEE

Enquête de conjoncture auprès des ménages (ECAMME) - 2012 - (2012) [fichier électronique], INSEE

Appendix A. Variance of the Test statistics (Bernardt et al., 2006)

If the realized variable π_{t+1} in period $t+1$ in relation to the agents j posterior (the median of j's posterior distribution over the realized variable) is given by,

$$\pi_{t+1} = \hat{\pi}^e_{j,t,t+1} + \epsilon_{j,t+1} \tag{A.1}$$

where $\epsilon_{j,t+1} \sim G(.)$ and ϵ_{t+1} is independent and identically distributed over the period of measurement for the forecasted variable and $G(0) \equiv 0.5$ and if the forecast is unbiased, it holds that $\pi^e_{j,t,t+1} = \hat{\pi}^e_{j,t,t+1}$, and the overshooting/undershooting indicator is distributed binomially as,

$$\sum_t \delta_t^+ \sim \mathcal{B}(\sum_t \gamma_t^+, G(0)) \quad \sum_t \delta_t^- \sim \mathcal{B}(\sum_t \gamma_t^-, 1 - G(0)) \tag{A.2}$$

This means the test statistic $S(z_t^-, z_t^+)$ is asymptotically normal distributed as

$$S \sim \mathcal{N}(0.5, \frac{1}{16}\left[\frac{1}{\sum_t \gamma_t^+} + \frac{1}{\sum_t \gamma_t^-}\right]) \tag{A.3}$$

since the mean of $S(z_t^-, z_t^+)$ is

$$\frac{1}{2}\left[P(\pi^e_{j,t,t+1} > \pi_{t+1}) + P(\pi^e_{j,t,t+1} < \pi_{t+1})\right] = 0.5[(1 - G(0)) + G(0)] = 0.5 \tag{A.4}$$

and Variance corresponds to

$$\begin{aligned}
&Var(S(z_t^-, z_t^+)) = Var(\frac{\sum_t \delta_t^+}{\sum_t \gamma_t^+}) + Var(\frac{\sum_t \delta_t^-}{\sum_t \gamma_t^-}) \\
&= \frac{1}{4(\sum_t \gamma_t^+)^2} Var(\sum_t \delta_t^+) + \frac{1}{4(\sum_t \gamma_t^-)^2} Var(\sum_t \delta_t^-) \\
&= \frac{1}{4(\sum_t \gamma_t^+)^2} G(0)(1 - G(0))(\sum_t \gamma_t^+) + \frac{1}{4(\sum_t \gamma_t^-)^2} G(0)(1 - G(0))(\sum_t \gamma_t^-) \\
&= \frac{G(0)(1 - G(0))}{4}\left[\frac{1}{\sum_t \gamma_t^+} + \frac{1}{\sum_t \gamma_t^-}\right]
\end{aligned} \tag{A.5}$$

Accordingly in the case of a commonly unforecasted shock to the forecasted variable, ω_{t+1} [7], $\sum_t \delta_t^+ \sim \mathcal{B}(\sum_t \gamma_t^+, G(-\omega_t))$ and $\sum_t \delta_t^- \sim \mathcal{B}(\sum_t \gamma_t^-, 1 - G(-\omega))$, which however leaves the mean, given the forecast is unbiased $\pi^e_{j,t,t+1} = \hat{\pi}^e_{j,t,t+1}$) of $S(.)$,

$$\begin{aligned}
&\frac{1}{2}\left[P(\pi^e_{j,t,t+1} > \pi_{t+1}) + P(\pi^e_{j,t,t+1} < \pi_{t+1})\right] \\
&= \frac{1}{2}\left[P(\hat{\pi}^e_{j,t,t+1} > \hat{\pi}^e_{j,t,t+1} + \omega_{t+1} + \epsilon_{j,t+1}) + P(\hat{\pi}^e_{j,t,t+1} < \hat{\pi}^e_{j,t,t+1} + \omega_{t+1} + \epsilon_{j,t+1})\right] \\
&= \frac{1}{2}\left[P(\epsilon_{j,t+1} < -\omega_{t+1}) + P(\epsilon_{j,t+1} > -\omega_{t+1})\right] \\
&0.5[G(-\omega_{t+1}) + (1 - G(-\omega_{t+1}))] = 0.5
\end{aligned}$$

[7]thus with commonly unforecasted shock ω_{t+1}, $\pi_{t+1} = \hat{\pi}^e_{j,t,t+1} + \omega_{t+1} + \epsilon_{j,t+1}$

, unaltered. The variance on the other hand is,

$$\frac{G(-\omega_{t+1})(1-G(-\omega_{t+1}))}{4}\left[\frac{1}{\sum_t \gamma_t^+}+\frac{1}{\sum_t \gamma_t^-}\right] \leq \frac{1}{16}\left[\frac{1}{\sum_t \gamma_t^+}+\frac{1}{\sum_t \gamma_t^-}\right]$$

Therefore the variance is at a maximum if $\omega = 0$ as $G(-\omega_{t+1})(1-G(-\omega_{t+1})) \leq G(0)(1-G(0)) = \frac{1}{2}$, which means that cross-sectional correlation reduces the variance of the test statistic. The same argumentation can also be used to show that the test statistics is robust to the other two problems addressed above: a bias caused by optimism/pessimism, as well as measurement errors (Bernardt et al., 2006, pp. 664).

Appendix B. Robustness (Bernardt et al., 2006)

The herding test of Bernardt et al. (2006) is suited for the application to inflation expectations insofar as it is robust to systematic biases of respondents in their perceptions/expectations.

Commonly unforecasted shocks. With regard to inflation this could in example be a generally unanticipated rise of commodity (oil) prices due to the outbreak of an armed conflict which would eventually lead to a shortage of supply. In this case a general shortfall of forecasts with regard to the realized inflation, leading to the estimation of the conditional probability $P(\pi_{t+1} < \pi^e_{t,t+1}|\bar{\pi}^e_{t,t+1} < \pi^e_{t,t+1}, \pi^e_{t,t+1} \neq \pi_{t+1}) > \frac{1}{2}$, and thus to the wrong conclusion that a herding-like bias was a reason for that. Consider a commonly unforecasted shock to the forecasted variable (here inflation) $\omega_{t+1} > 0$, be ϵ_{t+1} the idiosyncratic shock to inflation and G its cumulative distribution function. Despite a forecast is unbiased, $\hat{\pi}^e_{t,t+1} = \pi^e_{t,t+1}$, this would lead unconditionally to the conclusion of herding, $P(\pi_{t+1} < \pi^e_{t,t+1}|\bar{\pi}^e_{t,t+1} < \pi^e_{t,t+1}, \pi^e_{t,t+1} \neq \pi_{t+1}) < \frac{1}{2}$, since $P(\hat{\pi}^e_{t,t+1} + \omega_{t+1} + \epsilon_{t+1} < \pi^e_{t,t+1}) = P(\hat{\pi}^e_{t,t+1} + \omega_{t+1} + \epsilon_{t+1} < \hat{\pi}^e_{t,t+1}) = P(\epsilon_{t+1} < \omega_{t+1}) = G(-\omega_{t+1}) < \frac{1}{2}$. This is also true the other way around: with $\hat{\pi}^e_{t,t+1} = \pi^e_{t,t+1}$, $P(\hat{\pi}^e_{t,t+1} + \omega_{t+1} + \epsilon_{t+1} > \pi^e_{t,t+1}) = P(\hat{\pi}^e_{t,t+1} + \omega_{t+1} + \epsilon_{t+1} > \hat{\pi}^e_{t,t+1}) = P(\epsilon_{t+1} > -\omega_{t+1}) = 1 - G(-\omega_{t+1}) > \frac{1}{2}$, which implies $P(\pi_{t+1} < \pi^e_{t,t+1}|\bar{\pi}^e_{t,t+1} < \pi^e_{t,t+1}, \pi^e_{t,t+1} \neq \pi_{t+1}) > \frac{1}{2}$. A simple and effective solution, as pointed out by Bernardt et al. (2006), could be to use the average, as ω_t can be assumed to have offsetting effects. Under the assumption of unbiasedness $\hat{\pi}^e_{t,t+1} = \pi^e_{t,t+1}$, the test statistic, despite of the commonly unforecasted shock, would still yield 0.5, as $0.5[G(-\omega_{t+1}) + (1 - G(-\omega_{t+1}))] = \frac{1}{2}$.

Optimism/Pessimism. The same would be true if a certain degree of optimism or pessimism distorts the forecasts. The phenomenon discussed by Bernardt et al. (2006) is that forecasts further in the past tend to be more optimistic, while forecasts nearer to the disclosure of the realized value, as more information becomes available, tend to be more pessimistic. Such an effect can be modeled by introducing a bias α_t which changes over time. In example one can assume that t days before the disclosure of the realized value an agent forecasts the α percentile of the forecasted value. The result could be a false conclusion with regard to the presence of herding as in the case of commonly unforecasted earning shocks. Analogously to the latter case, this problem can also be addressed by taking averages, since $\frac{1}{2}P(\pi_{t+1} < \pi^e_{t,t+1}|\bar{\pi}^e_{t,t+1} < \pi_{t+1}) + \frac{1}{2}P(\pi_{t+1} > \pi^e_{t,t+1}|\bar{\pi}^e_{t,t+1} > \pi_{t+1}) = \frac{1}{2}[\alpha_t + (1 - \alpha_t)] = \frac{1}{2}$ rendering the conclusion if herding was present or not, unaltered.

Measurement errors. The problem of measurement errors might become relevant, as the value targeted by the agent in his forecast and its measurement differ (a different perception of price changes due to different consumption behavior). As mentioned above an individual might base its estimation of price changes on a mix of products compared to the basket of goods used by the statistical office in the calculation of the official rate of inflation for a certain period of time: The agent targets π_{t+1} with $\pi^e_{t,t+1} = \hat{\pi}^e_{t,t+1}$ while the realized value π_{t+1} is measured as $\pi_{t+1} + \lambda_{t+1}$. Similarly, as $\frac{1}{2}P(\pi_{t+1} < \pi^e_{t,t+1}|\bar{\pi}^e_{t,t+1} < \pi^e_{t,t+1}) + \frac{1}{2}P(\pi_{t+1} > \pi^e_{t,t+1}|\bar{\pi}^e_{t,t+1} > \pi^e_{t,t+1}) = \frac{1}{2}[G(\lambda_{t+1}) + (1 - G(\lambda_{t+1}))] = \frac{1}{2}$, this effect can be offset by using the average.

Appendix C. Variables - coding

variable	type	question	code	description
age	numeric	age	-	-
childs	numeric	number of children	-	-
citysize	ordered var.		0	rural
			1	less 5000 inhabitants
			2	between 5,000 and 9,999
			3	between 10,000 and 19,999
			4	between 20,000 and 49,999
			5	between 50,000 and 99,999
			6	between 100,000 and 199,999
			7	between 200,000 and 1,999,999
			8	Paris metropolitan region
educ	ordered var.	education	1	primary education or less
			2	secondary education
			3	post secondary education
			4	tertiary education
econ_per	ordered var.	The general economic situation in France in the last 12 months	1	significantly worsened
			2	slightly worsened
			3	stayed the same
			4	improved a bit
			5	significantly improved
econ_exp	ordered var.	The general economic situation in France in the next 12 months	1	will worsen significantly
			2	will slightly worsen
			3	will stay the same
			4	will improve a bit
			5	will improve significantly
finan	categorical var.	What descibes best the financial situation of your household	1	enough income to save a sufficient amount
			2	enough income to save a bit
			3	sufficient income to cover expenses
			4	have to use reserves to cover expenses

			5	coverage of expenses only possible with borrowing
fracwork	numeric	fraction of persons in the household with a job	-	-
price_per	ordered var.	Do you thin that prices in the last 12 months have	1	decreased
			2	stagnated
			3	increased moderately
			4	increased strongly
price_exp		In comparison with the last 12 months how do you think the evolution of prices will be in the next 12 months	1	prices will go down
			2	prices will stay the same
			3	prices will increase with a smaller rate
			4	prices will increase with the same rate
			5	prices will increase with a faster rate
region	categorical var.	region of residence	11	le-de-France
			21	Champagne-Ardenne
			22	Picardie
			23	Haute-Normandie
			24	Centre
			25	Basse-Normandie
			26	Bourgogne
			31	Nord-Pas-de-Calais
			41	Lorraine
			42	Alsace
			43	Franche-Comt
			52	Pays de Loire
			53	Bretagne
			54	Poitou-Charentes
			72	Aquitaine
			73	Midi-Pyrnes
			74	Limousin
			82	Rhne-Alpes
			83	Auvergne
			91	Languedoc-Roussilon
			93	Provence-Alpes-Cte d'Azur
			94	Corse

revquart	ordered var.	income quartile	1 2 3 4	1st quartile 2nd quartile 3rd quartile 4th quartiles
sex	categorical var.		1 2	male female
workregime	categorical var.		1 2 0 or 9	full time part time don't know/no answer
spouse	categorical var.		1 2	yes no
occupation	categorical var.		1 2 3 4	yes no, unemployed no, retired no, inactive
birthyr	numeric	birthyear	-	-

Appendix D. Inflation perceptions and expectations for each cohort

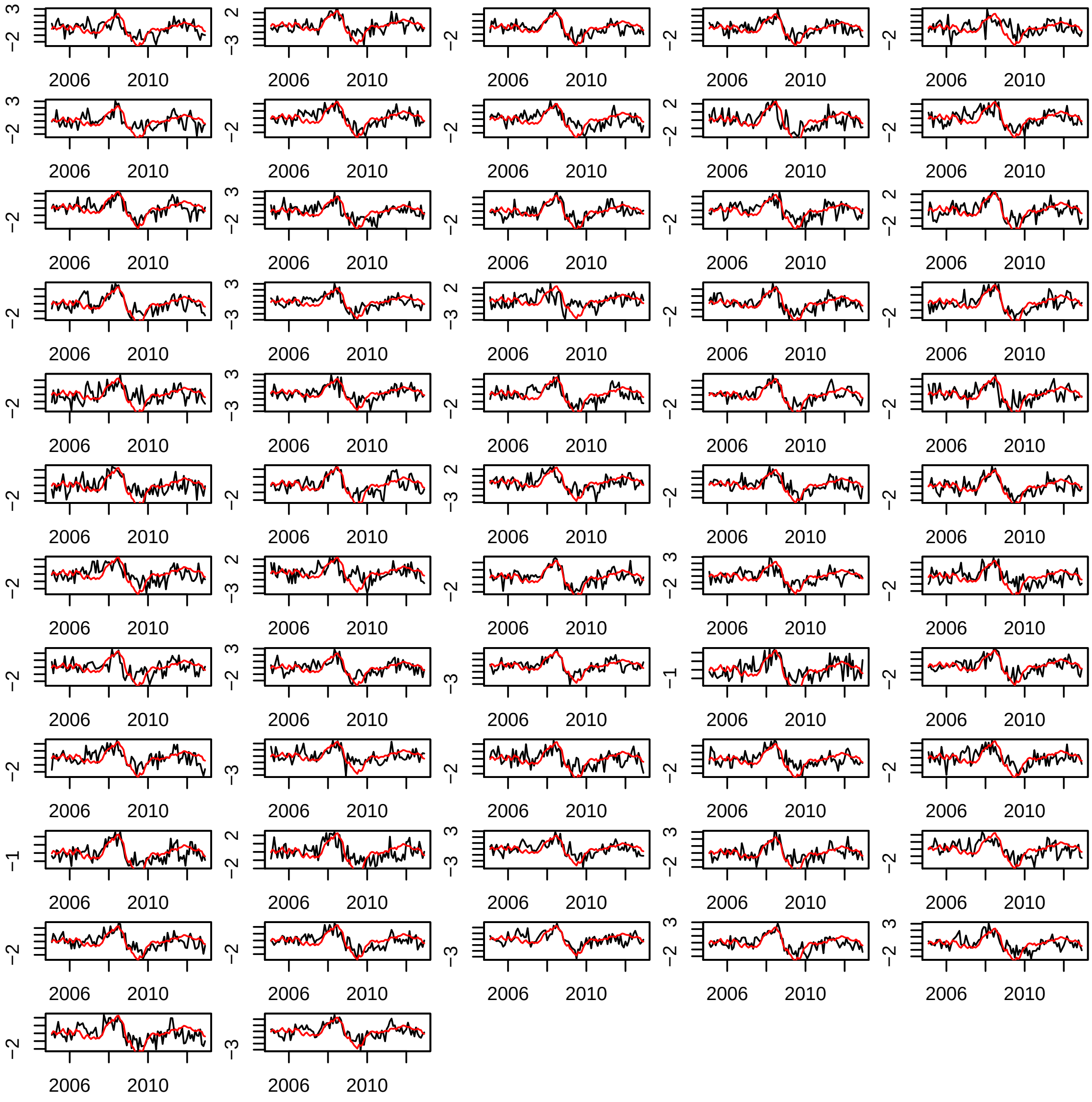

Figure D.9: Perceived inflation vs. actual inflation for each cohort

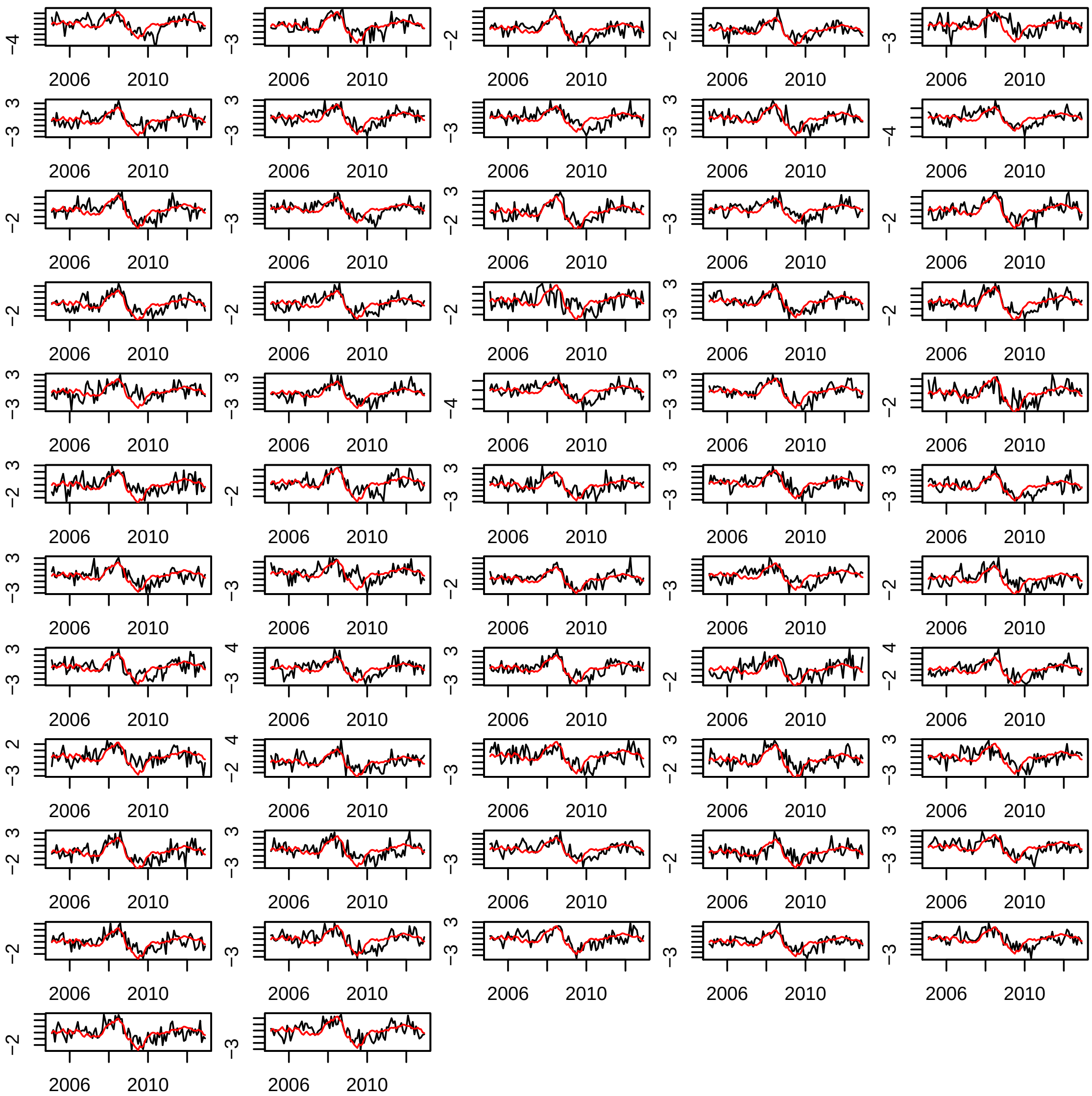

Figure D.10: Expected inflation vs. actual inflation for each cohort